IRE Transactions on INFORMATION THEORY

A Journal Devoted to the Theoretical and Experimental Aspects of Information Transmission, Processing and Utilization.

Volume IT-5 MARCH, 1959 Number 1
Published Quarterly

In This Issue

PUBLISHED BY THE
Professional Group on Information Theory

IRE Professional Group on Information Theory

The Professional Group on Information Theory is an organization, within the framework of the IRE, of members with principal professional interest in Information Theory. All members of the IRE are eligible for membership in the Group and will receive all Group publications upon payment of an annual fee of $3.00.

IRE TRANSACTIONS® on INFORMATION THEORY is published by the IRE for the Professional Group on Information Theory, at 1 East 79th Street, New York 21, N. Y. Responsibility for contents rests upon the authors and not upon the IRE, the Group, or its members. Price per copy: IRE-PGIT members, $1.00; IRE members, $1.50; nonmembers, $3.00.

INFORMATION THEORY

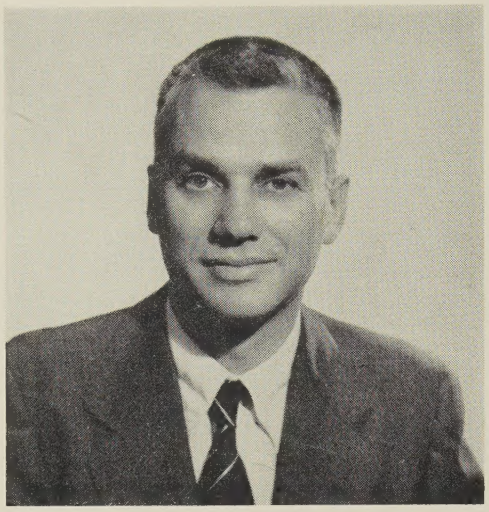
IRE Transactions on Information Theory

A Journal Devoted to the Theoretical and Experimental Aspects of Information Transmission, Processing and Utilization

Volume IT-5 March, 1959 Number 1

Published Quarterly

TABLE OF CONTENTS



J. L. Doob

J. L. Doob was born on February 27, 1910, in Cincinnati, Ohio. He attended Harvard University, Cambridge, Mass., from 1926 to 1932, receiving the B.A., M.A. and Ph.D. degrees. He was a National Research Fellow for the next two years, and was supported a further year by a grant from the Carnegie Corporation.

In September, 1945 he joined the faculty of the University of Illinois, Urbana, where he has remained, except for a period in Washington, D. C. working for the Navy during World War II.

Dr. Doob is a member of the Institute of Mathematical Statistics, the American Mathematical Society, and the National Academy of Sciences.

Editorial

As befits the role of an American mathematician in modern society, I have nothing practical to say about information theory. However the devotees of this theory may be interested in the reactions of an outsider who has followed some of its development.

In spite of all the suggestive work by Wiener, Shannon, and their successors, the main thing that strikes an outsider is that there are so few theoretical results. In fact almost every time a writer proves an assertion connecting the capacity of a channel with the entropy of a source, his paper P_n is succeeded by a paper P_{n+1} which, instead of generalizing or extending the results of P_n, is devoted to pointing out and correcting some defect or insufficiency in it. The paper P_{n+1}, in its turn, receives the same harsh treatment, and so on. Moreover, in this presumably convergent process of purging and purifying, the theorems become more and more attenuated and inapplicable as their hypotheses become more restrictive.

Even more extraordinary is the fact that this process of organizing what seems to be the very basis of the subject seems to have no effect whatever on its applications! Can it be that the existence of a mathematical basis is irrelevant, and that the basic principle is the very idea that there is a context in which the word "information" is accepted by general agreement and used in an intuitive way, and that no more is needed?

J. L. Doob

Correlation and Delay Line Attenuation*

MELVIN J. JACOBSON†

Summary—The effect of delay line attenuation on the output of a correlator is studied for the case where the attenuation in db to within a frequency independent loss varies linearly with delay and as the square root of frequency. It is shown how attenuation affects the output signal-noise ratio of the correlator, increasing it for some signal spectrum shapes and decreasing it for others. Upper bounds on the increase and decrease are computed and a sufficient condition on the input signal spectrum is established which, when satisfied, assures a degradation in signal-noise ratio. Also examined is the effect of a frequency independent gain inserted to compensate for the delay line attenuation. It is shown how this gain may be chosen in order to give uniform system output noise over all values of delay.

I. Notation

f = frequency
f_2 = upper frequency accepted by system
f_1 = lower frequency accepted by system
r = bandwidth ratio f_2/f_1
$s(f)$ = signal power density
$n(f)$ = noise power density
τ = relative delay in signal in the two system channels
ξ = inserted delay
S = average input signal power
N = average input noise power
k = delay line constant
$x_1, x_2 = k\xi f_1^{\frac{1}{2}}, k\xi f_2^{\frac{1}{2}}$

II. Introduction

It is well known that a system employing a cross-correlation device may be used to determine the presence of a signal in noise. A portion of a correlation system containing a delay line is shown in Fig. 1.[1] In channel 1, $s(t)$ is the signal voltage, $n_1(t)$ is an undesired noise voltage, and t is a measure of time. In channel 2, an undesired noise voltage $n_2(t)$ is present together with the desired signal voltage $s(t + \tau)$. Note that the signal in channel 1 is delayed relative to that in channel 2 by the time τ. The information in channel 2 is now delayed for a time ξ after which the voltages in the two channels enter the correlator where they are multiplied together and averaged over a finite time interval to give the output of the system. When ξ takes on the value τ, the signal at both inputs to the correlator will be the same and maximum signal cross-correlation will result. The delay line may be switched to channel 1 to determine the

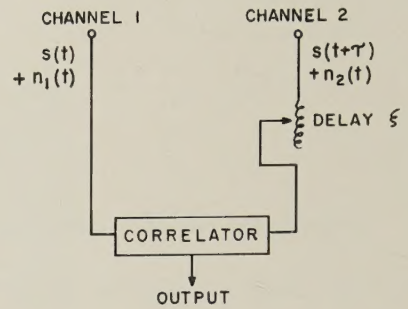


Fig. 1—Portion of a delay line correlation system.

presence of a signal which is delayed in channel 2 relative to channel 1.

It is the purpose of this paper to investigate how attenuation in the delay line influences the effectiveness of the correlation system in determining the presence of a signal in noise. In particular, we shall be concerned with an examination of the output signal-noise ratio and the output noise of the system described above when the delay line attenuation in db, to within a frequency independent loss, varies linearly with delay and as the square root of frequency. Such an attenuation will result if the input frequencies are in the low megacycle region and if the delay line makes use of uniformly dissipative lumped constant delay networks having air core solenoid coils designed for minimum loss.[2-4] Of course, this restriction on the input frequencies to the delay line and correlator does not require that the signal and noise be in the low-megacycle region initially.

In the following sections it will be assumed that the noise in the two system channels have the same power spectra but do not correlate. Further, it will be assumed that the signal and noise do not correlate and that both the signal and noise possess the ergodic and gaussian properties.

III. The Ideal System

When a correlation system contains a loss-free delay mechanism, the signal and noise power spectra in the two system channels are the same, and the representation for the mean and variance of the output are well known. In particular, the mean system output is given by

* Manuscript received by the PGIT, March 24, 1958. The work leading to this paper was supported in part by Office of Naval Research Contract Nonr-591(09).

† Dept. of Mathematics, Rensselaer Polytechnic Inst., Troy, N. Y.

[1] For another analysis of a similar system, see P. E. Green, Jr., "The output signal-to-noise ratio of correlation detectors," IRE Trans. on Information Theory, vol. IT-3, pp. 10-18; March, 1957.

[2] J. F. Blackburn, "Components Handbook," M. I. T. Rad Lab. Ser., McGraw-Hill Book Co., New York, N. Y., ch. 6; 1949

[3] S. Ramo and J. R. Whinnery, "Fields and Waves in Moder Radio," John Wiley and Sons, Inc., New York, N. Y., sec. 6.1 1944.

[4] A. T. Starr, "Electric Circuits and Wave Filters," Pitma Publ. Corp., New York, N. Y., pp. 124-127; 1944.

[5] J. J. Faran, Jr. and R. Hills, Jr., "Correlators for Signa Reception," Acoust. Res. Lab., Harvard University, Cambridg Mass., Tech. Memo. 27; 1952.

$$R_{12}(\xi - \tau) = \int_{f_1}^{f_2} s(f) \cos 2\pi f(\xi - \tau)\, df \qquad (1)$$

d the variance is given by

$$\sim \frac{1}{2RC} \int_{-\infty}^{\infty} [R_{11}(x)R_{22}(x) + R_{12}(x + \xi - \tau)R_{21}(x - \xi + \tau)]\, dx \qquad (2)$$

hen the correlator contains an RC averaging network d $RC(f_2 - f_1) \gg 1$. The quantities R_{11} and R_{22} are the itocorrelations of the outputs of channels 1 and 2 spectively, and R_{12} and R_{21} are cross correlations of the itputs of the two channels. For the particular case of reat interest where the input signal power is much less an the input noise power, we may take

$$R_{11}(x) = R_{22}(x) \sim \int_{f_1}^{f_2} n(f) \cos 2\pi f x\, df \qquad (3)$$

d neglect cross-correlation terms in (2) to get

$$\sigma^2 \sim \frac{1}{2RC} \int_{-\infty}^{\infty} \left[\int_{f_1}^{f_2} n(f) \cos 2\pi f x\, df \right]^2 dx. \qquad (4)$$

we also suppose that the noise power spectrum is flat er $f_1 < f < f_2$, then

$$n(f) = \frac{N}{f_2 - f_1} \qquad (5)$$

d the variance reduces further to

$$\sigma^2 \sim \frac{N^2}{4RC(f_2 - f_1)}. \qquad (6)$$

the noise spectrum is not flat, one can always equalize make it so.

One accepted measure of the quality of this type of stem is the maximum value of the output signal-noise wer ratio, defined as the ratio of the square of (1) to) with $\xi - \tau$ equated to zero. When $\xi - \tau = 0$,

$$R_{12}(0) = \int_{f_1}^{f_2} s(f)\, df = S \qquad (7)$$

d the output signal-noise ratio is given by

$$\left(\frac{S}{N}\right)_{out} = 4RC(f_2 - f_1)\left(\frac{S}{N}\right)^2. \qquad (8)$$

IV. Inclusion of Attenuation

We shall now examine the effect of delay line attenuion on correlator output signal-noise ratio when the equency range of the input to the delay line is in the w megacycle region. Of course, this does not necessarily strict the frequencies initially accepted by the system lie in this range. With the type of delay line described Section I we may assume that, except for a flat (frequency independent) loss, the attenuation in the line in is equal to $8.69k\xi f^{1/2}$ where k is a delay line constant. hen the attenuation in nepers is equal to $k\xi f^{1/2}$.

We begin by assuming that the noise voltage v_i at the put to the delay line may be expanded in a Fourier series over the time interval $0 \leq t < T$ to give

$$v_i(t) = \frac{a_0}{2} + \sum_{m=1}^{\infty} a_m \cos \frac{2\pi m t}{T} + b_m \sin \frac{2\pi m t}{T} \qquad (9)$$

where the a_m and b_m are independent variables distributed normally with zero mean and the same variance, and m/T is the frequency of the m'th sinusoidal component of the noise.[6] If this noise voltage is now passed through the delay line, the output to within a flat loss is given by

$$v_0(t) = \frac{a_0}{2} + \sum_{m=1}^{\infty} \left[a_m \cos \frac{2\pi m(t - \xi)}{T} + b_m \sin \frac{2\pi m(t - \xi)}{T} \right] \exp [-k\xi(m/T)^{1/2}] \qquad (10)$$

for $\xi \leq t < T + \xi$. If we now form the product of $v_0(t)$ and $v_0(t - x)$, average over time, and let T become infinite, it is found that the autocorrelation of the noise in the delayed channel is given by

$$\int_{f_1}^{f_2} n(f) \exp (-2k\xi f^{1/2}) \cos 2\pi f x\, df. \qquad (11)$$

The autocorrelation of the noise in the undelayed channel is given by (3) and, analogous to (4), the variance of the system output is given by

$$\sigma^2 \sim \frac{1}{2RC} \int_{-\infty}^{\infty} \left[\int_{f_1}^{f_2} n(f) \cos 2\pi f x\, df \right] \cdot \left[\int_{f_1}^{f_2} n(f) \exp (-2k\xi f^{1/2}) \cos 2\pi f x\, df \right] dx \qquad (12)$$

to within a flat loss. When the input noise spectrum is flat, the first inner integral is easily evaluated and (12) may be written as

$$\sigma^2 \sim \frac{N^2}{4\pi RC(f_2 - f_1)^2} \int_{f_1}^{f_2} \exp (-2k\xi f^{1/2}) \cdot \int_{-\infty}^{\infty} \frac{\sin 2\pi f_2 x - \sin 2\pi f_1 x}{x} \cos 2\pi f x\, dx\, df. \qquad (13)$$

The integral over x is equal to π[7] so that

$$\sigma^2 \sim \frac{N^2}{4RC(f_2 - f_1)^2} \int_{f_1}^{f_2} \exp (-2k\xi f^{1/2})\, df. \qquad (14)$$

In a similar fashion, it can be shown that the mean system output is given by

$$R_{12}(0) = \int_{f_1}^{f_2} s(f) \exp (-k\xi f^{1/2})\, df \qquad (15)$$

when $\xi - \tau = 0$ and the output signal-noise ratio may be written as

$$\left(\frac{S}{N}\right)_{out} = 4RC(f_2 - f_1)^2 \left(\frac{S}{N}\right)^2 \frac{\left[\int_{f_1}^{f_2} \frac{s(f)}{S} \exp (-k\xi f^{1/2})\, df \right]^2}{\int_{f_1}^{f_2} \exp (-2k\xi f^{1/2})\, df}. \qquad (16)$$

[6] S. O. Rice, "Mathematical analysis of random noise," *Bell Sys. Tech. J.*, vol. 23, pp. 282–332; July, 1944. (Sec. 2.3.)

[7] D. B. DeHaan, "Nouvelles Tables D'Intégrales Définies," G. E. Strechert and Co., New York, N. Y.; 1939. (Table 151.)

It can be seen at this point that a frequency independent loss in the delay line has no effect upon output signal-noise ratio, since the appropriate factors, if included in (14) and (15), would cancel in (16).

In the two sections to follow, we shall be interested in the ratio of the output signal-noise ratio for the attenuation case to that for the ideal case. From (8) and (16), this ratio is given by

$$D = \frac{(f_2 - f_1)\left[\int_{f_1}^{f_2} \frac{s(f)}{S} \exp(-k\xi f^{1/2})\, df\right]^2}{\int_{f_1}^{f_2} \exp(-2k\xi f^{1/2})\, df}. \tag{17}$$

A value of D less than one will represent a degradation in output signal-noise ratio resulting from delay line attenuation, while a value of D greater than one will represent an improvement in signal-noise ratio.

V. Constant Slope Signal Spectra

When a signal spectrum has a constant slope of 3a db per octave, the power density may be written as

$$s(f) = Kf^a \tag{18}$$

where K must satisfy

$$K = \frac{S}{\int_{f_1}^{f_2} f^a\, df} \tag{19}$$

if the signal is to have average power S in the band $f_1 < f < f_2$. If (18) and (19) are substituted into (17) and if $f_2 - f_1$ is given an integral representation, then we may write

$$D(a) = \frac{\int_{x_1}^{x_2} x\, dx\left[\int_{x_1}^{x_2} x^{2a+1}e^{-x}\, dx\right]^2}{\int_{x_1}^{x_2} xe^{-2x}\, dx\left[\int_{x_1}^{x_2} x^{2a+1}\, dx\right]^2}. \tag{20}$$

When the signal spectrum is flat, $a = 0$ and this expression reduces to

$$D(0) = \frac{\left[\int_{x_1}^{x_2} xe^{-x}\, dx\right]^2}{\int_{x_1}^{x_2} xe^{-2x}\, dx \int_{x_1}^{x_2} x\, dx}. \tag{21}$$

Now, by Schwarz's inequality for integrals,[8]

$$\left[\int_{x_1}^{x_2} xe^{-x}\, dx\right]^2 = \left[\int_{x_1}^{x_2} x^{1/2}(x^{1/2}e^{-x})\, dx\right]^2 < \int_{x_1}^{x_2} x\, dx \int_{x_1}^{x_2} xe^{-2x}\, dx \tag{22}$$

so that $D(0) < 1$ for $x_2 > x_1$. Thus, a delay line attenuation of the type being considered here will give rise to a degradation in output signal-noise ratio for all frequency bands and inserted delays when the signal spectrum is flat over the band $f_1 < f < f_2$.

We shall now demonstrate that $D(a) < D(b)$ when $a > b$. That is,

$$\frac{\int_{x_1}^{x_2} x\, dx\left[\int_{x_1}^{x_2} x^{2a+1}e^{-x}\, dx\right]^2}{\int_{x_1}^{x_2} xe^{-2x}\, dx\left[\int_{x_1}^{x_2} x^{2a+1}\, dx\right]^2} < \frac{\int_{x_1}^{x_2} x\, dx\left[\int_{x_1}^{x_2} x^{2b+1}e^{-x}\, dx\right]^2}{\int_{x_1}^{x_2} xe^{-2x}\, dx\left[\int_{x_1}^{x_2} x^{2b+1}\, dx\right]^2}. \tag{23}$$

This desired inequality may be written as

$$\int_{x_1}^{x_2} x^{2a+1}e^{-x}\, dx \int_{x_1}^{x_2} x^{2b+1}\, dx - \int_{x_1}^{x_2} x^{2b+1}e^{-x}\, dx \int_{x_1}^{x_2} x^{2a+1}\, dx < 0. \tag{24}$$

Now, after appropriate manipulation (see Appendix), the quantity on the left may be written as

$$\frac{1}{2}\int_{x_1}^{x_2}\int_{x_1}^{x_2} (xy)^{2b+1}[x^{2(a-b)} - y^{2(a-b)}][e^{-x} - e^{-y}]\, dx\, dy. \tag{25}$$

This is an integral in the first quadrant of the x, y plane over a square having a diagonal coincident with the line $y = x$. The integrand is negative throughout this square, except on the diagonal where it is equal to zero. Hence the value of the integral is negative and the desired inequality is proved. The implication of this result is that degradation in output signal-noise ratio increases as the slope of a signal spectrum having constant db per octave slope increases.

If $D(a)$ is expanded in a Taylor series about $x_2 = 0$, it is found that

$$D(a) = 1 + \left[\frac{4(r^{3/2} - 1)}{3r^{1/2}(r - 1)} - \frac{4(a + 1)(r^{(2a+3)/2} - 1)}{(2a + 3)r^{1/2}(r^{a+1} - 1)}\right]x_2 + O(x_2^2) \tag{26}$$

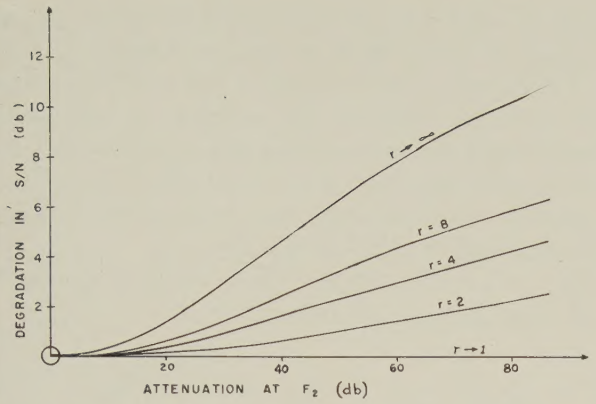
for $a \neq -1$. The coefficient of x_2 is negative for $a > 0$ and positive for $a < 0$ so that $D < 1$ for $a > 0$ and $D > 1$ for $a < 0$. Thus, the output signal-noise ratio for a decreasing (negative slope) signal spectrum is actually increased by delay line attenuation when $8.69x_2$—the attenuation in db at the upper frequency of the acceptance band—is sufficiently small.

For decreasing as well as increasing spectra, there is a steadily increasing degradation in output signal-noise ratio as x_2 increases since, for x_2 large,

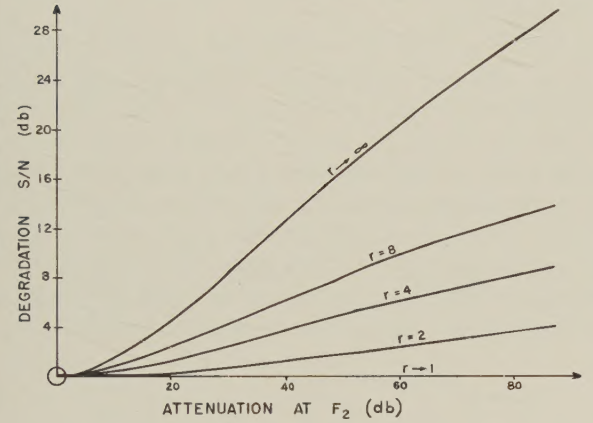
$$D(a) \sim \frac{4(a + 1)^2(r - 1)r^{1/2}}{(r^{a+1} - 1)^2 x_2} \tag{27}$$

for $a \neq -1$ and

[8] G. H. Hardy, J. E. Littlewood, and G. Polya, "Inequalities," Cambridge University Press, Cambridge, Mass., p. 132; 1934.



g. 2—Degradation in output signal-noise ratio vs delay line attenuation at upper frequency. Flat signal spectrum.



g. 3—Degradation in output signal-noise ratio vs delay line ttenuation at upper frequency. 3 db per octave signal spectrum.

$$D(-1) \sim \frac{4(r-1)r^{1/2}}{(\log r)^2 x_2} \tag{28}$$

r $a = -1$.

The degradation D is plotted in db vs $8.69x_2$, the tenuation in db at the upper frequency f_2, in Figs. 2, and 4. These figures correspond to $a = 0$, 1, and -1 signal spectra having constant slopes of 0, 3, and -3 per octave. In each figure D is plotted for $r = 2$, 4, and these values corresponding to bandwidths of 1, 2, and octaves. Curves are also plotted for the limiting cases $\rightarrow 1$ and $r \rightarrow \infty$. Note that the curves of Fig. 4 show a gative degradation or an improvement in output signal-ise ratio relative to the loss-free case when x_2 is suf-iently small.

VI. More General Spectra

Since attenuation is least for small frequencies and eatest for large frequencies, minimum degradation will cur when the signal is represented by a line spectrum $f = f_1$ while maximum degradation will result when e signal spectrum is a line at $f = f_2$. Letting f approach in (15) gives

$$R_{12}(0) = Se^{-k\xi f_1^{1/2}}. \tag{29}$$

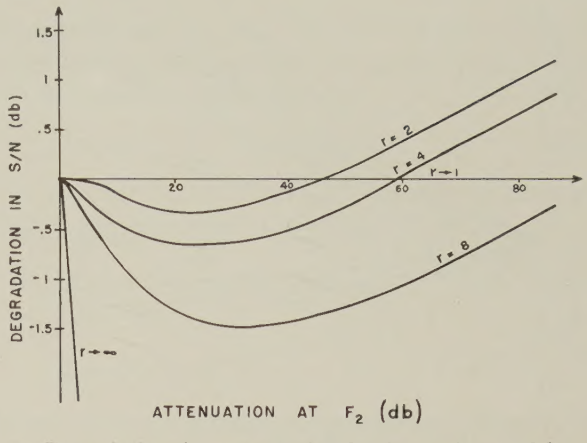


Fig. 4—Degradation in output signal-noise ratio vs delay line attenuation at upper frequency. −3 db per octave signal spectrum.

The variance given by (14) and the output signal-noise ratio of the ideal system given by (8) remain unchanged. From (8), (14), and (29) the minimum degradation is given by

$$D_{\max} = \frac{2(x_2^2 - x_1^2)e^{-2x_1}}{e^{-2x_1}(2x_1 + 1) - e^{-2x_2}(2x_2 + 1)}. \tag{30}$$

Similarly, the maximum degradation is given by

$$D_{\min} = \frac{2(x_2^2 - x_1^2)e^{-2x_2}}{e^{-2x_1}(2x_1 + 1) - e^{-2x_2}(2x_2 + 1)}. \tag{31}$$

The maximum and minimum degradation are plotted in db as a function of $8.69x_2$ in Fig. 5. Note that, for fixed r, the curves of Figs. 2, 3, and 4 are bounded by those of Fig. 5.

From (30) and (31) it can be seen that the largest possible variation in the output signal-noise ratio of the delay line correlation system is given by

$$10 \log_{10} \exp 2(x_2 - x_1) = 8.69(1 - 1/r^{1/2})x_2 \text{ db}. \tag{32}$$

This variation appears graphically in Fig. 6 as a function of x_2.

In addition to the above results, it is also possible to establish a sufficient condition on the signal spectrum which, when satisfied, assures a degradation in output signal-noise ratio relative to that of the ideal system. From (17) it can be seen that D will be less than one if

$$(f_2 - f_1)\left[\int_{f_1}^{f_2} \frac{s(f)}{S} \exp(-k\xi f^{1/2})\, df\right]^2 < \int_{f_1}^{f_2} \exp(-2k\xi f^{1/2})\, df. \tag{33}$$

If $s(f)$ is a continuous function of f, then (33) will be satisfied if

$$(f_2 - f_1)\left[\frac{s_{\max}}{S} \int_{f_1}^{f_2} \exp(-k\xi f^{1/2})\, df\right]^2 < \int_{f_1}^{f_2} \exp(-2k\xi f^{1/2})\, df \tag{34}$$

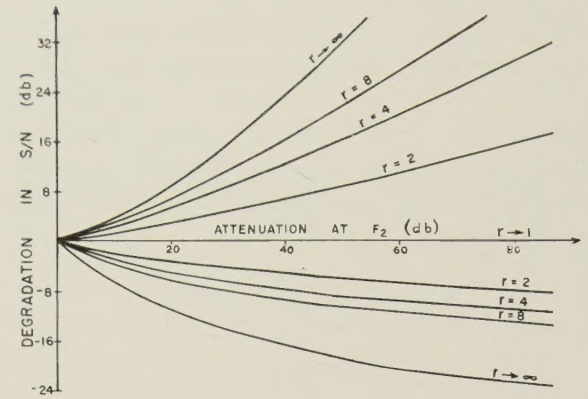


Fig. 5—Maximum and minimum degradation in output signal-noise ratio vs delay line attenuation at upper frequency.

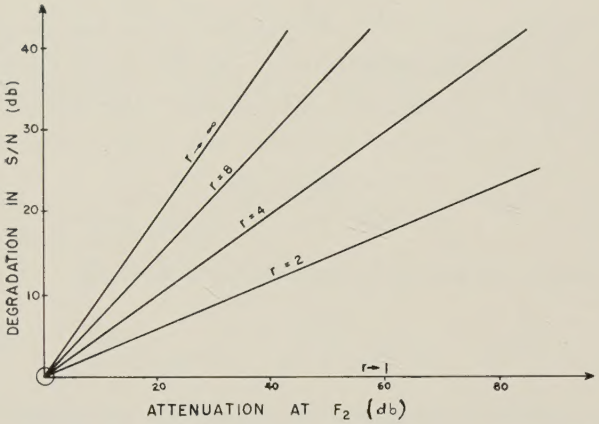


Fig. 6—Maximum possible variation in output signal-noise ratio vs delay line attenuation at upper frequency.

where $s_{\max}$ is the largest value of the signal power density between f_1 and f_2. If we now note that

$$S = (f_2 - f_1)s_{av} \tag{35}$$

where s_{av} is the average value of $s(f)$ between f_1 and f_2 then (34) may be written as

$$\frac{s_{\max}}{s_{av}} < \frac{\left[\int_{x_1}^{x_2} x\,dx\right]^{1/2}\left[\int_{x_1}^{x_2} xe^{-2x}\,dx\right]^{1/2}}{\int_{x_1}^{x_2} xe^{-2x}\,dx} = D^{-1/2}(0) \tag{36}$$

where $D(0)$ is the degradation in output signal-noise ratio for a flat signal spectrum and is plotted in db in Fig. 2.

VII. Output Noise

In Section IV an expression was developed for the variance of the output of the delay line correlation system. Eq. (14) indicates that the variance or output noise is a function of the inserted delay and, in fact, decreases as ξ increases. In searching for a signal in noise it would be desirable to have a uniform output noise for all values of inserted delay, and it is the purpose of this section to indicate a method which will provide such a result. In particular, we shall assume that in addition to a gain inserted to offset the flat loss of the delay line, a compensating gain is added which is equal to $8.69k\xi f_0^{\frac{1}{2}}$ db where f_0 is a frequency lying between f_1 and f_2. This compensation could be obtained by inserting an appropriate variable attenuator and amplifier following the delay line. It will be noted that this gain will not affect the results of the previous sections, since the inclusion of frequency independent loss or gain will not alter (17).

With the insertion of the gain described above, the output noise of the correlator is given by

$$\sigma^2 \sim \frac{N^2}{4RC(f_2 - f_1)^2}\int_{f_1}^{f_2} \exp 2k\xi(f_0^{1/2} - f^{1/2})\,df. \tag{37}$$

If we now require that σ^2 not only be constant but equal to the output noise of the ideal system, then from (6) and (37) we have that

$$\frac{N}{(f_2 - f_1)}\int_{f_1}^{f_2} \exp 2k\xi(f_0^{1/2} - f^{1/2})\,df = N. \tag{38}$$

The right side of this equation is the noise power at the input to the delay line and the left side is the noise power at the output of the amplifier. Hence, the above requirements on the correlator output noise are analogous to the requirement that there be no loss in noise power in the delayed channel of the system. Evaluation of (38) gives

$$e^{2x_0} = \frac{2(x_2^2 - x_1^2)}{(2x_1 + 1)e^{-2x_1} - (2x_2 + 1)e^{-2x_2}} \tag{39}$$

where $x_0 = k\xi f_0^{\frac{1}{2}}$. For fixed values of the frequency range and inserted delay, the above expression gives the frequency f_0 at which unity gain should be required in order to have no loss in noise power in the delayed channel. This frequency is a function of ξ and therefore changes as the value of the inserted delay is changed. The relation between $8.69x_0$ and $8.69x_2$ as given by (39) appears graphically in Fig. 7 for various values of r. The figure may also be interpreted as giving the required amount of compensating gain $8.69x_0$ db.

VIII. Conclusion

In this paper, some effects of delay line attenuation on the output of a correlation system have been examined for the case where the input to the delay line is in the low megacycle region and the line contains uniformly dissipative lumped constant delay networks having air core solenoid coils designed for minimum loss. In this case the attenuation in db to within a flat loss, varies linearly with delay and as the square root of frequency. The output of this system has been compared to that of an ideal system in which there is no attenuation in the delayed channel.

It has been shown that the output signal-noise ratio of the delay line system may be larger or smaller than that of the ideal system depending upon the shape of the input signal spectrum. Upper and lower bounds on the output ratio of the delay line system have been found. In addition

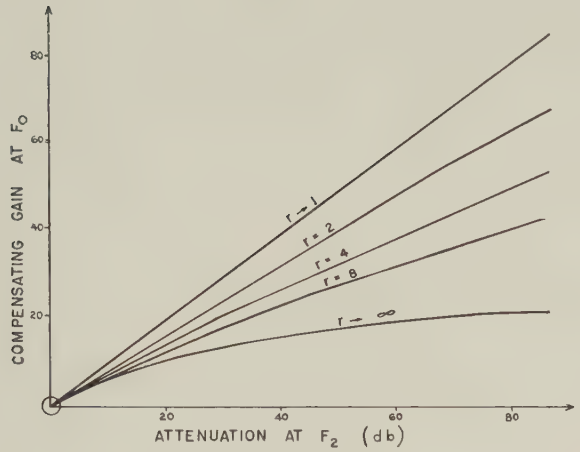


Fig. 7—Compensating gain vs delay line attenuation at upper frequency for uniform output noise.

condition on the signal spectrum has been established which, when satisfied, assures a degradation in output signal-noise ratio resulting from delay line attenuation. These results are independent of a frequency independent compensating gain which might be added in the delayed channel. It has also been shown that delay line attenuation affects the output noise of the system, causing it to decrease as the amount of delay is increased. It is desirable to have constant output noise over delay, and it has been demonstrated how this may be accomplished by inserting an appropriate compensating gain in the dealyed channel which is independent of frequency.

Appendix

The quantity on the left of (24) may be written in terms of double integrals as

$$\int_{x_1}^{x_2}\int_{x_1}^{x_2} x^{2a+1}e^{-x}y^{2b+1}\,dx\,dy - \int_{x_1}^{x_2}\int_{x_1}^{x_2} y^{2b+1}e^{-y}x^{2a+1}\,dx\,dy$$

$$= \int_{x_1}^{x_2}\int_{x_1}^{x_2} x^{2a+1}y^{2b+1}(e^{-x} - e^{-y})\,dx\,dy$$

$$= \frac{1}{2}\int_{x_1}^{x_2}\int_{x_1}^{x_2} x^{2a+1}y^{2b+1}(e^{-x} - e^{-y})\,dx\,dy$$

$$+ \frac{1}{2}\int_{x_1}^{x_2}\int_{x_1}^{x_2} x^{2a+1}y^{2b+1}(e^{-x} - e^{-y})\,dx\,dy.$$

Interchanging x and y in the second integral and combining the two resulting double integrals gives

$$\frac{1}{2}\int_{x_1}^{x_2}\int_{x_1}^{x_2} (xy)^{2b+1}[x^{2(a-b)} - y^{2(a-b)}][e^{-x} - e^{-y}]\,dx\,dy. \quad (25)$$

As described in Section V, the value of this integral is necessarily negative since the integrand is negative throughout the region of integration, except along the line $y = x$ where its value is zero.

On the First Probability of Detection by a Radar Receiver System*

W. M. STONE†, R. L. BROCK†, AND K. J. HAMMERLE‡

Summary—**Expressions for the detection probabilities associated with the output of filter-square law detector-filter radar receivers are presented for practical filter systems and with a slowly varying Rayleigh distributed signal amplitude.**

KAC AND Siegert[1] and Emerson[2] have discussed the first probability of detection of a signal in the presence of noise using a filter-square law detector-filter system. The present authors[3] have extended the theory in two ways: 1) the signal is assumed to be randomly modulated, and 2) the filters employed are physically realizable in the sense of zero response before zero time. This paper presents the principal results achieved.

It is considered that a discrete sampling of a continuous signal in the presence of noise is performed at the system output. The criterion of performance is measured by the probability, $P(x_0, \gamma, z)$, that a sampled value of the system output shall exceed a given bias level, where x_0 is the normalized bias level, γ is related to the ratio of band-

* Manuscript received by the PGIT, June 11, 1958.
† Boeing Airplane Co., Seattle, Wash., and Oregon State College, Corvallis, Ore.
‡ Boeing Airplane Co., Seattle, Wash.
[1] M. Kac and A. J. F. Siegert, "On the theory of noise in radio receivers with square law detectors," *J. Appl. Phys.*, vol. 18, pp. 383–397; April, 1947.
[2] R. C. Emerson, "First probability densities for receivers with square law detectors," *J. Appl. Phys.*, vol. 24, pp. 1168–1176; September, 1953.
[3] W. M. Stone and R. L. Brock, "On the Probability of Detection with a Postdetection Filter," Boeing Airplane Company Document D-16921; 1955.

widths of the two filters, and z is the expected value of the signal to noise power ratio.

Two tacit assumptions[2,4] are usually made in the first probability theory for signal detection: 1) the threshold device gets only one "look" at the system output when the signal is present and 2) the signal has been present for so long a time that its buildup in the filters can be ignored and modulation such as that due to antenna scanning need not be considered. These assumptions are really contradictory. A better approach would be to consider the actual signal envelope shape and to determine the probability that the bias level will be exceeded at least once during the presence of the fleeting signal. A preliminary study along these lines has been accomplished by the present writers.

The first of these two assumptions will usually lead to no significant error when a properly designed system is considered. In such a system the time constant of the second filter will be roughly equal to the signal duration so that, in general, the threshold device will get only one "look" at the signal as the radar beam scans across the target.

To compensate for the error introduced by the second of these assumptions, the signal to noise ratio may be adjusted. When the second filter bandwidth and the input signal modulation envelope are known, the time function representing the signal presented to the threshold in the absence of noise is readily determined. The ratio of the peak value of this function to the peak value that would be reached with an audio filter of negligibly short time constant can be calculated. This ratio, together with the "law" of the detector, enables one to make the appropriate correction. A second correction to allow for a signal which is off-center in frequency is readily incorporated into the theory.

The basic formulations of the receiver system are the same as those of Kac and Siegert and Emerson. Let the first filter possess the frequency response function,

$$F_1(\omega) = \left[1 + j\frac{\omega - \omega_0}{\omega_1}\right]^{-1} + \left[1 + j\frac{\omega + \omega_0}{\omega_1}\right]^{-1}, \qquad (1)$$

$$\omega_0 = k\omega_1, \qquad k \gg 1,$$

and let the second filter possess the response function,

$$F_2(\omega) = \left[1 + j\frac{\omega}{\omega_2}\right]^{-1}. \qquad (2)$$

In dimensionless variables the system kernel of the pertinent homogeneous integral equation may be readily determined as

$$g(x, y) = \frac{2\omega_1^2}{2\gamma - 1}\cos 2k\gamma(x - y) \cdot \begin{cases} e^{-2\gamma(x+y)}[e^{2(2\gamma-1)x} - 1], & 0 \le x \le y, \\ e^{-2\gamma(x+y)}[e^{2(2\gamma-1)y} - 1], & 0 \le y \le x, \end{cases} \qquad \gamma = \omega_1/\omega_2. \qquad (3)$$

[4] D. Middleton, "Statistical criteria for the detection of pulsed carriers in noise," *J. Appl. Phys.*, vol. 24, pp. 371–391; April, 1953.

In principle the system kernel is developable in a bilinear expansion of the set of orthonormal functions,

$$h_n^{C,S}(x) = \frac{2e^{-x}J_{2\gamma-1}(r_n e^{-x})}{J_{2\gamma}(r_n)} \cdot \begin{cases} \cos 2k\gamma x, \\ \sin 2k\gamma x, \end{cases} \quad 0 \le x < \infty, \quad J_{2\gamma-1}(r_n) = 0 \qquad (4)$$

but the orthonormal property holds only for relatively large values of $k\gamma$. Eq. (4) is equivalent to (7.18) and (7.19) of Kac and Siegert.

For greater generality the signal is assumed to be sinusoidal but of frequency not necessarily equal to the center frequency ω_0 of the first filter,

$$S(t) = A\cos\omega t. \qquad (5)$$

Since $\omega_0 \gg \omega_1$ it may be specified that

$$\alpha = |\omega - \omega_0|/\omega_1 \qquad (6)$$

is of order unity while $(\omega + \omega_0)/\omega_1 \gg 1$. A set of functions of immediate interest is

$$F_{2\gamma-1}^{C,S}(\alpha, r_n) = \int_0^\infty e^{-x}J_{2\gamma-1}(r_n e^{-x}) \begin{cases} \cos 2\gamma\alpha x \\ \sin 2\gamma\alpha x \end{cases} dx, \qquad (7)$$

which is implicit in (5.18) of Kac and Siegert.

Without further ado the Laplace transform of the probability density function associated with the output of the system takes the form

$$\phi(s, \gamma, y) = \phi(s, \gamma, 0) \cdot \exp\left[-y^2\gamma\sum_{k=1}^{\infty}(-1)^{k-1}(8\gamma s)^k B_k(\alpha, \gamma)\right], \qquad (8)$$

where y is the normalized signal amplitude, $\phi(s, \gamma, 0)$ is the Laplace transform for the noise only case, and the B_k are defined by

$$B_k(\alpha, \gamma) = \sum_{n=1}^{\infty}\frac{F_{2\gamma-1}^{C^2}(\alpha, r_n) + F_{2\gamma-1}^{S^2}(\alpha, r_n)}{r_n^{2k+4}J_{2\gamma}^2(r_n)}, \qquad k = 1, 2, \cdots. \qquad (9)$$

At least the first four of the B_k are obtainable in closed form. The probability that a sampled value of output, noise only input, exceeds a given bias level is obtained as

$$P(x_0, \gamma, 0) = \frac{1}{\Gamma(2\gamma)}\sum_{n=1}^{\infty}\frac{(r_n/2)^{2\gamma-2}e^{-r_n^2x_0/8\gamma}}{J_{2\gamma}(r_n)}, \qquad (10)$$

which has been tabulated in a Boeing document[3] for $x_0 = 1.5(0.1)3.5$ and $\gamma = 0(1)40$. Eq. (10) is equivalent to (57) of Siegert.[5] For signal amplitude a constant the only possibility of inverting the Laplace transform of (8) is to resort to an Edgeworth series. But it makes physical sense to assume that y is a Rayleigh distributed random variable. Such a distribution is defined by

[5] A. J. F. Siegert, "A systematic approach to a class of problems in the theory of noise and other random phenomena—Part II, examples," IRE TRANS. ON INFORMATION THEORY, vol. IT-3, pp. 38–43; March, 1957.

$$f(y) = \begin{cases} \frac{1}{z} y e^{-y^2/2z}, & y \geq 0, \\ 0, & y < 0, \end{cases} \tag{11}$$

here z is the average signal to noise power ratio. The ective average signal to noise power ratio is

$$z_1 = \frac{z}{1 + \alpha^2}, \tag{12}$$

d, finally, the probability of detection for relatively rge γ becomes

$$P(x_0, \gamma, z_1) = \frac{(2\gamma/z_1)^{(2\gamma-1)/2} e^{-x_0/z_1}}{\Gamma(2\gamma) J_{2\gamma-1}[2(2\gamma/z_1)^{1/2}]} \tag{13}$$

$$= e^{-x_0/z_1}\left[1 - \frac{1}{z_1} + \frac{\gamma}{z_1^2(2\gamma + 1)} - \frac{\gamma^2}{3z_1^3(2\gamma + 1)(\gamma + 1)} + \cdots\right]^{-1}.$$

he series form would be useful only for $z_1 > 1$.

In similar fashion a system with a double-tuned LC st filter,

$$(\omega) = \left[1 + \left(\frac{\omega - \omega_0}{\omega_1}\right)^2 + 2bj\frac{\omega - \omega_0}{\omega_1}\right]^{-1} + \left[1 - \left(\frac{\omega + \omega_0}{\omega_1}\right)^2 + 2bj\frac{\omega + \omega_0}{\omega_1}\right]^{-1}, \tag{14}$$

and a second filter specified by (2) leads to

$$P(x_0, \gamma, z_1) = e^{-x_0/z_1}\left[1 - \frac{1}{z_1} + \frac{\gamma^2(4b\gamma + 1)}{z_1^2(4\gamma^2 + 4b\gamma + 1)(2b\gamma + 1)} - \cdots\right]^{-1},$$

$$\gamma = \omega_1/\omega_2, \tag{15}$$

a result not significantly different from (13). The probability of detection for a Rayleigh distributed signal amplitude for Gaussian filters is developed with little difficulty. It is recognized that a Gaussian filter is not physically realizable in the sense previously defined. The moments for all three systems are also given in the referenced document.[3]

Large γ implies that the post-detection smoothing is very good, so that for a given signal strength the output of the second filter is proportional to the power through the first filter, with negligible variance about this value. If z_1 is assumed large the power present would be almost all signal power. Since the signal amplitude follows a Rayleigh distribution the signal power would be exponentially distributed, as is indicated by the leading terms of the series in (33) and (15).

The authors wish to express their appreciation for the many remarks and suggestions of the referee.

Full Decodable Code-Word Sets*

M. P. SCHÜTZENBERGER† AND R. S. MARCUS‡

Summary—This paper considers further how the decodability condition imposes restrictions on a set of code words. A generating function is defined that describes the composition of the code words. The relation between the generating function and a "full" set of code words is found. This relation shows that the sum of arbitrary probabilities associated with the words of a full set must be one. A full set of code words is one to which no code word can be added and still keep the set decodable. It is also shown that a full set is "completable." For a completable set of code words any string of symbols can be made into a sentence by adding a suitable prefix and a suffix.

Introduction

SEVERAL authors have considered the restrictions that are imposed on the set of code words by the decodability condition.[1-5] (A code-word set is decodable if no string of symbols can be broken up into code words in more than one way.) Most of the results thus far have had to do with the *lengths* of the code words. This paper includes some conclusions relating to the more detailed *composition* of the code words.

It is important to consider the composition of the code words, as well as their lengths, when the symbols are not of the same cost. For example, in the Morse code the dot is shorter in time duration than the dash. The less costly dot, therefore, should be used more frequently for efficiency of information transmission.

In particular, this paper defines a generating function that describes the composition of the code words. The relation between this function and a "full" set of code words is found. A full set of code words is one to which no code words can be added and still keep the set decodable. It is also shown that a full set is "completable." For a completable set of code words, any string of symbols can be made into a sentence by adding a suitable prefix and a suffix.

* Manuscript received by the PGIT, July 25, 1958. This work was supported in part by the U. S. Army (Signal Corps), the U. S. Air Force (Office of Sci. Res., Air Res. and Dev. Com.), and the U. S. Navy (Office of Naval Research).

† Faculté des Sciences de Poitiers, France; formerly with Res. Lab. of Electronics, Mass. Inst. Tech., Cambridge, Mass.

‡ Res. Lab. of Electronics, Mass. Inst. Tech., Cambridge, Mass.

[1] A. A. Sardinas and G. W. Patterson, "A necessary and sufficient condition for unique decomposition of coded messages," 1953 IRE National Convention Record, pt. 8, pp. 104–108.

[2] B. Mandelbrot, "On recurrent noise limiting coding," *Proc. Symp. on Information Networks*, New York, N. Y.; 1954.

[3] B. McMillan, "Two inequalities implied by unique decipherability," IRE Trans. on Information Theory, vol. IT-2, pp. 115–116; December, 1956.

[4] M. P. Schützenberger, "On an application of semi-group methods to some problems in coding," IRE Trans. on Information Theory, vol. IT-2, pp. 47–60; September, 1956.

[5] R. S. Marcus, "Discrete noiseless coding," S. M. thesis, Dept. Elec. Eng., M. I. T., Cambridge, Mass.; January, 1957.

Statement of the Problem

Let us consider an information-carrying channel with D symbols, d_j, $j = 1, \cdots, D$. For any given string of symbols, s, we write $|s|_j \equiv$ the number of occurrences of symbol d_j in s, and $|s| \equiv$ the total number of symbols in s. Thus, $|s| = \sum_j |s|_j$. A *code word*, w_k, is a particular s. The *code-word set*, P_0, is a set of M code words. *Sentences* are strings of words and they form the infinite set $P = \{P_0\}$. It is always understood that the lengths of the code words are bounded. Without this hypothesis, the conclusions are somewhat different.

It is convenient to associate with the set $\{d_j\}$ an arbitrary set of probabilities, p_j ($\sum p_j = 1$, $p_j > 0$, $j = 1, \cdots, D$). Then we write $Pr(s) = \prod_j p_j^{|s|_j}$. We may now define the *generating function of the words*, $\phi_{P_0}(t)$:

$$\phi_{P_0}(t) = \sum_k Pr\,(w_k) t^{|w_k|} = \sum_{i=1}^{n_m} a_i t^i, \tag{1}$$

where

$$a_i = \sum_{|w_k| = i} Pr\,(w_k)$$

$$n_m = \max\,\{|\,w_k\,|\}.$$

Similarly, we define the *generating function of the sentences*, $\Phi_P(t)$:

$$\Phi_P(t) = \sum_{s \varepsilon P} \nu(s)\, Pr\,(s) t^{|s|} = \sum_n A_n t^n, \tag{2}$$

where

$$A_n = \sum_{|s| = n} \nu(s)\, Pr\,(s)$$

$\nu(s)$ = number of decompositions of s into words.

A code-word set, P_0, is then uniquely decodable or, let us say, just *decodable* (d), if $\nu(s) = 1$ for all s in P. (Of course, $\nu(s) = 0$, if s is not in P.) P_0 is said to be *full* (F) if no word can be added to P_0 to form a code-word set that is decodable. P_0 is said to be *completable* (C) if any string, s, can be made to fit in P by adding some suitable prefix and suffix. (Symbolically, we write: P_0 is C if $\forall\ s\ \exists\ x$ and y ʒ $xsy\ \varepsilon\ P$.)[6]

The four theorems that will be presented show that the four following statements are equivalent for decodable code-word sets:

I. P_0 is full.

II. P_0 is completable.

III. $\phi_{P_0}(1) = 1$ for some particular p_j set. (3)

IV. $\phi_{P_0}(1) \equiv 1$ for all p_j sets.

[6] The symbols xsy denote the string x, followed by the string s, followed by the string y. Here x and y may vary for different s's. $\forall$ means for all; $\exists$ means there exists; ʒ means with the property that; ε means belonging to.

Theorems

Theorem I: If P_0 is C, then $\phi_{P_0}(1) = 1$.

Method of Proof: Since the sentences are defined recursively, the A_n are given by a difference equation and are the sums of roots to the nth power, as shown in section 1). For P_0 completable, we show that the A_n cannot become vanishingly small, as shown in sections 2)–4). But for P_0 decodable, the A_n cannot become larger than one. Thus the root of minimum modulus, the real root, must be one.

Proof:

1)
$$A_n = \sum_{i=1}^{n_m} B_i T_i^{-n}, \tag{4}$$

where T_i are the roots of $\phi_{P_0}(t) = 1$
B_i are constants.

Eq. (4) is true, since A_n is given by

$$A_n = \sum_{i=1}^{n_m} a_i A_{n-i}. \tag{5}$$

The solution of the difference (5) is given by

$$A_n = \sum_{i=1}^{n_m} B_i \rho_i^n, \tag{6}$$

where ρ_i are roots of $\rho^{n_m} - a_1\rho^{n_m-1} - \cdots - a_{n_m} = 0$
B_i are constants.
Letting $T = \rho^{-1}$, we have

$$T^{-n_m} - a_1 T^{1-n_m} - a_2 T^{2-n_m} - \cdots - a_{n_m} = 0$$
$$= 1 - a_1 T - a_2 T^2 - \cdots - a_{n_m} T^{n_m}$$
$$= 1 - \phi_{P_0}(T).$$

This proves (4).

2) If P_0 is C, then the number of symbols in any prefix and suffix that is needed to make s in P is bounded. More specifically, we have

$$|x| + |y| \leq L = 2n_{m-1}. \tag{7}$$

This is obviously true, since if $|x| > n_m$ we could break up x into words and a string x' with the property that $|x'| < n_m$. This x' could serve as a suitable prefix; similarly for y.

3) If P_0 is C, then $\sum_{n=\alpha}^{\alpha+L} A_n > C_1 > 0$; (for any α). (8)

To prove this, let u_i be the $D^\alpha s \ni |s| = \alpha$. (See Fig. 1.)

Let u_i' be one $xu_iy \in P \ni |x| + |y| \leq L$. Hence,

$$\alpha \leq |u_i'| \leq \alpha + L.$$

Some of the u_i' may be the same but we can pick a set of distinct u_i', say v_i, with the property that each u_i can be expanded into at least one v_j. Let $u_{i,j}$ be the set of u_i that can be expanded into a given v_j.

Let

$$\sum_{u_i \in u_{i,j}} Pr(u_i) \equiv Pr(u_{i,j}).$$

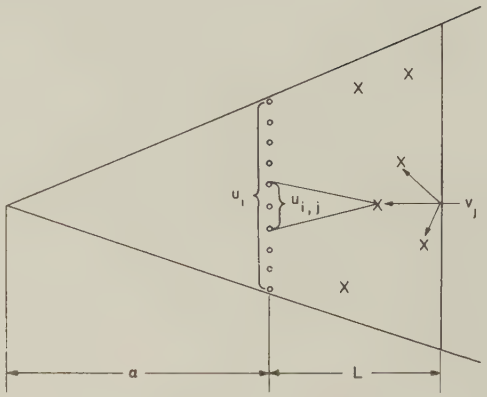


Fig. 1—Abstraction from code-word tree.

Then

$$\sum_j Pr(u_{i,j}) \geq \sum_i Pr(u_i) = 1.$$

Now from each $u_{i,j}$ pick the u_i (call it w_j) with the maximum $Pr(u_i)$. The maximum number of u_i in any $u_{i,j}$ is $|v_j| - \alpha + 1$. An upper bound on this number is $(\alpha + L) - \alpha + 1 = L + 1$. Thus $Pr(w_j) \geq [1/(L+1)]\, Pr(u_{i,j})$. But $Pr(v_j) = Pr(x)\, Pr(w_j)\, Pr(y) \geq p_{\min}^L\, Pr(w_j)$, where $p_{\min} = \min\{p_i\}$. Hence,

$$\sum_{n=\alpha}^{\alpha+L} A_n \geq \sum_j Pr(v_j) \geq p_{\min}^L \sum_j Pr(w_j)$$
$$\geq \frac{p_{\min}^L}{L+1} \sum_j Pr(u_{i,j}) \geq \frac{p_{\min}^L}{L+1} \equiv C_1 > 0$$

This proves (8).

4) Hence, $\lim_{n\to\infty} A_n > C_1/(L+1) \equiv C_2 > 0$ (if the limit exists).

5) Hence, $|T_1| \leq 1$, where $|T_1|$ is the minimum modulus. If $|T_1| > 1$, $A_n \to 0$ as $n \to \infty$.

6) A_n must be bounded. If A_n were not bounded, then $\nu(s)$ would be greater than one for some s because

$$A_n = \sum_{|s|=n} \nu(s)\, Pr(s) \quad \text{and} \quad \sum_{|s|=n} Pr(s) = 1.$$

This would mean that P_0 is not d, contrary to the hypothesis that P_0 is C.

7) Hence, $|T_1| \geq 1$. Otherwise A_n would be unbounded.

8) Since all the coefficients of $\phi_{P_0}(t)$ are positive, $\phi_{P_0}(t)$ is monotonic and $\phi_{P_0}(t) = 1$ has one real root, and no other root has a modulus smaller than this.[7]

9) Hence, $|T_1| = T_1 = 1$.

10) Hence, $\phi_{P_0}(1) = 1$ (and this is true for all p_i sets).

Theorem II: If $\phi_{P_0}(1) = 1$ and P_0 is d, then P_0 is F.

Proof: If we add a word to P_0 to give P_0', then $\phi_{P'}{}_0(1) > 1$, and T_1', the real root of $\phi_{P'}{}_0(t) = 1$, is less than one. But by Theorem I, section 6), and Theorem I, section 7), this implies that P_0' is not d. Thus P_0 is F.

[7] The fact that the real root has the minimum modulus follows from Cauchy's theorem. Cf. Morris Marden, "The Geometry of the Zeros of a Polynomial in a Complex Variable," Mathematical Surveys No. III, American Mathematical Society, New York, N. Y.; 1949. See especially Theorem (27.1), p. 95.

Theorem III: If P_0 is d and $\phi_{P_0}(1) = 1$ for a given p_i, then P_0 is C.

Proof: Suppose P_0 is not C. Then $\exists\ s_0 \ni \forall x, y\ xs_0y \notin P$. Since no s with s_0 as a prefix is in P, those strings in that part of the tree that "grows" from s_0 can be eliminated as possible s in P. This means that for $n \geq |s_0|$,

$$A_n \leq 1 - Pr(s_0).$$

Of those strings that do not begin with s_0, we can eliminate that fraction whose second $|s_0|$ symbols are s_0. Thus

$$A_n \leq [1 - Pr(s_0)]^2 \quad \text{for} \quad n \geq 2 \mid s_0 \mid.$$

Similarly,

$$A_n \leq [1 - Pr(s_0)]^m \quad \text{for} \quad n \geq m \mid s_0 \mid.$$

Hence, $A_n \to 0$ as $n \to \infty$. But for given p_i, $T_1 = 1$ and $A_n > C_3 > 0$ for some $n > N$ for any N. Hence, we have a contradiction and P_0 is C.

Theorem IV: If P_0 is F, then P_0 is C.

Method of Proof: Assuming that P_0 is not completable, we consider the string, u, which cannot be completed. If we add u as a word to P_0, we obtain a new set, $\bar{P}_0$, which cannot be decodable. We then show that this implies that u has the same string of symbols in its beginning as at its end, as shown in section 14). But this leads to a contradiction.

Proof:

1) Assume that P_0 is F but not C.

2) Hence, $\exists\ u \ni \forall\ x, y\ xuy \notin P = \{P_0\}$.

3) Consider $\bar{P}_0 = P_0 \cup u$ and $\bar{P} = \{\bar{P}_0\}$.

4) Since $\bar{P}_0$ is not decodable, $\exists\ v$ with two decompositions in $\bar{P}$.

5) Choose v as a minimal doubly decomposable string (minimal d.d. string); that is, a string that cannot remain d.d. if any symbols are removed from its beginning and/or end.

6) Since P_0 is d and $\bar{P}_0$ is not, one of the decompositions of v must contain u as a word. Thus $v = x_1uy_1$, where $x_1, y_1 \in \bar{P}$.

7) Since u is not completable in P, $v \notin P$.

8) But $v \in \bar{P}$.

9) Hence, the second decomposition of v also contains u, *i.e.*, $v = x_2uy_2$.

10) Assume that $|x_1| \leq |x_2|$. If this is not so, reverse designations.

11) $|x_2| \neq |x_1|$. If $|x_2| = |x_1|$, then $x_2 = x_1$, and for v to be d.d. either $x_1 = x_2$ is d.d. or $y_1 = y_2$ is d.d., contrary to the hypothesis that v is a minimal d.d. string.

12) Hence, $|x_1| < |x_2|$.

13) Let us so choose the second decomposition that $|x_2| < |x_1| + |u|$. (See Fig. 2.)

Otherwise, x_2 contains u and must be decomposed as $x_2 = x_3uy_3$ by the same reasoning that led to section 9). Thus we could have chosen to consider the first u as the word u in the second decomposition of v.

14) Thus $u = x_4u_2 = u_2y_4$. (See Fig. 3.)

15) We can find (as we shall show) a u' for which the equation of section 14) cannot be satisfied. Hence, the assumption that P_0 is F, but not C, which leads to this conclusion, is false and the theorem is proved.

16) To find u' we consider two cases that cover all the possibilities.

Case 1): $u = a^{|u|}$.

Case 2): $u = a^kby_5$; $0 < k < |u|$, $0 \leq |y_5|$.

We have arbitrarily called the first symbol in u "a" and the first symbol in u which is not a, if such a symbol exists, "b".

17) For case 1), let $u' = ub = a^{|u|}b$.

Clearly, u' cannot satisfy

$$u' = x_5w = wy_6; \qquad |\ y_6\ | > 0,$$

since w must start with a and end with b.

18) For case 2), let $u' = ub^{|u|}$.

$|w| \leq |u|$ is clearly impossible, for then w would have to start with "a" but consist only of b's.

But if $|w| > |u|$, we can write

$$w = x_6ab^{r+|u|},$$

where $0 \leq |x_6|$; $0 \leq r < |u|$.

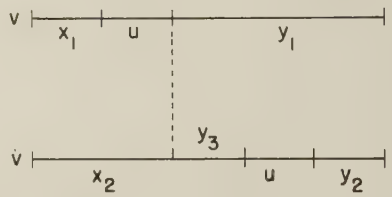


Fig. 2—Grouping of symbols in the string v.

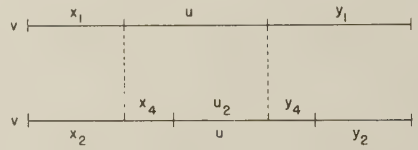


Fig. 3—Grouping of symbols in the string v.

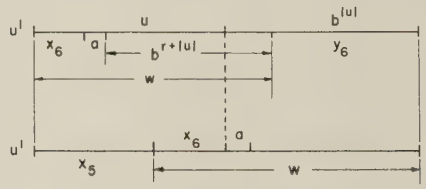


Fig. 4—Grouping of symbols in the string u'.

Then, as is apparent from Fig. 4, $u' = x_5w$ requires that the "a" in question occur in a position that must be a "b" from the fact that $u' = wy_6$. This contradiction shows that the given u' for case 2) does not satisfy section (14). Thus section 15) is proved and Theorem IV, in turn, is proved.

Conclusion

The four theorems, taken together, show the logical equivalence of the four properties of the statements of equation 3), as is indicated in Fig. 5.

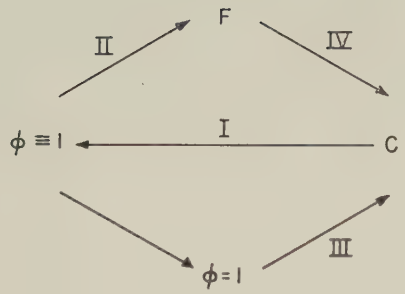


Fig. 5—Diagram showing the relations of the four theorems.

Sections 1)–5) of Theorem I then show that the probilities associated with a full code-word set must sum at st to one. Sections 6) and 7) of Theorem I show that s sum must be no more than one if the code is decodable; at is, $\phi_{P_0}(1) \leq 1$ if P_0 is d. It can easily be shown that s inequality leads to the generalized Kraft[8] inequality

$$\sum_{k=1}^{M} 2^{-q_k} \leq 1,$$

where q_k is the normalized cost of word w_k.

Our discussion shows that the equality sign holds only when P_0 is full. This inequality was obtained by Marcus[5] by extending Mandelbrot's proof[2] for the equal-cost case. Mandelbrot used Shannon's Fundamental Theorem for Discrete Noiseless Channels[9] and pointed out that a similar inequality had been obtained previously by Szilard. McMillan[3] obtained a proof in the equal-cost case without using information-theory concepts. Note that the proofs of this paper are also independent of the Shannon theorem.

For the equal-cost case, the normalized cost is just $q_k = n_k \log D$, with $n_k = |w_k|$. Thus the inequality reads:

$$\sum_{k=1}^{M} D^{-n_k} \leq 1.$$

[8] L. G. Kraft, "A device for quantizing, grouping and coding plitude modulated pulses," S. M. thesis, Dept. Elec. Eng., I. T. Cambridge, Mass.; 1949.

[9] C. E. Shannon, "A mathematical theory of communication," *Bell Sys. Tech. J.*, vol. 27, pp. 379–423; July, 1948.

On a Property of Wiener Filters*

MOSHE ZAKAI†

Summary—Let $Y(\omega,\alpha)$ be the Wiener filter designed to yield output which is the least-square approximation to $s(t + \alpha)$ ere $s(t)$ is the desired signal input. Let $Y_L(\omega)$ be the Wiener er designed to yield an output which is the least-square approxiation to some linear operation L on the desired input signal. e following simple relationship has been shown to hold between $\omega, \alpha)$ and $Y_L(\omega)$. If $s(t)$ is the desired input signal and $L_\alpha[s(t + \alpha)$ he desired output, where L_α is some linear operation with respect α, then $Y_L(\omega) = L_\alpha[Y(\omega, \alpha)]$.

WE consider the well-known Wiener problem (Fig. 1). Let $s(t)$ be the signal process and let $Y(\omega, \alpha)$ be the physically realizable transfer nction that yields an output which is the least-meanuare approximation to $s(t + \alpha)$. A generalization of s problem is to ask for the filter $Y_L(\omega)$ which yields the st-mean-square approximation to $f(t)$, where $f(t)$ is e result of some linear operation on $s(t)$. For example,

$$f(t) = \frac{ds(t)}{dt} \quad \text{or} \quad f(t) = \int_t^{t+\alpha_1} s(\tau)\, d\tau.$$

As is well known, $Y_L(\omega)$ can be found by solving the problem for the given operation.[1,2] It is the purpose of this paper to show that once we know $Y(\omega, \alpha)$ for all real α, and for a given set of spectral densities $G_s(\omega)$, $G_n(\omega)$, $G_{sn}(\omega)$, $G_{ns}(\omega)$, we can find $Y_L(\omega)$ directly from $Y(\omega, \alpha)$ without any further references to the spectral densities from which $Y(\omega, \alpha)$ was derived.

Furthermore, we can find $Y_L(\omega)$ from $Y(\omega, \alpha)$ by the following operation. Because of the linearity of the operation we can write

$$f(t) = L_\alpha[s(t + \alpha)] \tag{1}$$

where $f(t)$ is the desired output, $s(t)$ is the signal input and L is the linear operation with respect to the variable α leading from $s(t + \alpha)$ to $f(t)$. Then $Y_L(\omega)$ is given by

$$Y_L(\omega) = L_\alpha[Y(\omega, \alpha)]. \tag{2}$$

* Manuscript received by the PGIT, March 24, 1958. This is ondensed version of Tech. Rep. T-8/133, written while the author s at Columbia University Electronics Res. Labs., New York, Y., July 15, 1958. The research reported was sponsored by the ctronics Res. Dir., AF Cambridge Res. Cntr., Air Res. and v. Com., under Contract AF 19(604)-1572.

† Scientific Dept., Ministry of Defence, Israel.

[1] H. S. Tsien, "Engineering Cybernetics," McGraw-Hill Book Co., Inc., New York, N. Y.; 1954.

[2] L. A. Zadeh and J. R. Ragazzini, "An extension of Wiener's theory of prediction," *J. Appl. Phys.*, vol. 21, pp. 645-655; July, 1950. (See ex. 1, p. 653.)

A similar result holds for the impulsive responses. For example, let $f(t) = ds(t)/dt$.
Then

$$f(t) = \frac{ds(t+\alpha)}{d\alpha}\bigg|_{\alpha=0}; \qquad L_1 = \left(\frac{d}{d\alpha}\right)_{\alpha=0}.$$

Therefore

$$Y_{L_1}(\omega) = \left(\frac{dY(\omega,\alpha)}{d\alpha}\right)_{\alpha=0}.$$

As a second example we consider the operation of integration

$$f(t) = \int_t^{t+\alpha} s(\tau)\, d\tau;$$

then

$$L_2[s(t+\alpha)] = \int_0^{\alpha_1} s(t+\alpha)\, d\alpha$$

and

$$Y_{L_2} = \int_0^{\alpha_1} Y(\omega,\alpha)\, d\alpha.$$

The proof given below can easily be extended to show that (2) remains true for the Zadeh-Ragazzini finite memory filters with stationary inputs[2] and for multiple-input filters.[3]

We consider now the mean-square error of the system of Fig. 1. If the desired operation L on the signal is characterized by a transfer function $Q(\omega)$ then the expression for the mean-square system error expressed in the frequency domain is given by[1]

$$\overline{\epsilon^2(t)} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \{G_s \mid Y - Q \mid^2 + G_{ns}Y^*(Y-Q) + G_{sn}Y(Y^* - Q^*) + G_n \mid Y \mid^2\}\, d\omega.$$

This expression can also be written as

$$\overline{\epsilon^2(t)} = \frac{1}{2\pi}\int_{-\infty}^{\infty} \begin{bmatrix} Y-Q \\ Y \end{bmatrix}^{*T} \begin{bmatrix} G_s & G_{sn} \\ G_{ns} & G_n \end{bmatrix} \begin{bmatrix} Y-Q \\ Y \end{bmatrix} d\omega. \quad (3)$$

We now set up the following Hilbert space interpretation of the optimization problem.[4] The following is an extension of the Hilbert space representation for the noiseless prediction problem[5] to include noisy inputs. We consider the linear vector space of vectors of the form

$$F(\omega) = \begin{bmatrix} \sum_{\nu=0}^{N} a_\nu e^{i\alpha_\nu\omega} \\ \sum_{\mu=0}^{M} a_\mu e^{i\alpha_\mu\omega} \end{bmatrix}. \quad (4)$$

The scalar product $[K_1(\omega), K_2(\omega)]$ is defined as

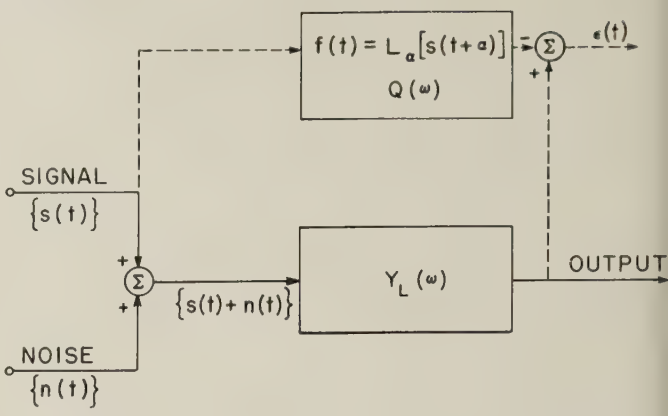


Fig. 1

$$(K_1, K_2) = \frac{1}{2\pi}\int_{-\infty}^{\infty} (K_1^*(\omega))^T \begin{bmatrix} G_s & G_{sn} \\ G_{ns} & G_n \end{bmatrix} K_2(\omega)\, d\omega.$$

Completing the space by adding to it limits in the mean we obtain a Hilbert space H.

$Q(\omega)$ was defined to be the transfer function associated with the desired linear operation on the signal (Fig. 1). Therefore[6] $Q(\omega)$ is either of the form of $q(\omega)$ where

$$q(\omega) = \sum_{\nu=0}^{N} a_\nu e^{i\alpha_\nu\omega} \quad (\alpha_\nu \text{ real})$$

or $Q(\omega)$ is the limit in the mean of a sequence of such finite sums. The vector

$$\begin{bmatrix} Q(\omega) \\ 0 \end{bmatrix}$$

therefore belongs to the Hilbert space H defined above.

$Y(\omega)$ (Fig. 1) is the transfer function associated with a linear operation on the past of $\{s(t) + n(t)\}$. Therefore either $Y(\omega)$ has the form

$$Y(\omega) = \sum_{\nu=0}^{N} a_\nu e^{-i\alpha_\nu\omega}, \quad (\alpha_\nu \text{ real and non-negative}) \quad (5)$$

or else $Y(\omega)$ is a limit in the mean of such expressions. In other words, since the spectral density of $\{s(t) + n(t)\}$ is given by $(G_s + G_n + G_{sn} + G_{ns})$, either $Y(\omega)$ is of the form given above or else there exists a sequence $Y_n(\omega)$ with all $Y_n(\omega)$ of the form given above such that as $n \to \infty$

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} (G_s + G_n + G_{sn} + G_{ns}) \mid Y - Y_n \mid^2 d\omega \to 0,$$

but

$$(G_s + G_n + G_{sn} + G_{ns}) \mid Y \mid^2 = \begin{bmatrix} Y^* \\ Y^* \end{bmatrix} \begin{bmatrix} G_s & G_{sn} \\ G_{ns} & G_n \end{bmatrix} \begin{bmatrix} Y \\ Y \end{bmatrix}.$$

Hence the set of vectors $[Y, Y]^T$ constitute a subspace H_1 of H.

If we write the relation between the desired output $f(t)$

[3] N. Wiener, "Interpolation, Extrapolation and Smoothing of Stationary Time Series," John Wiley and Sons, Inc., New York, N. Y.; 1950.
[4] It was suggested to us by Prof. L. Zadeh that by sacrificing rigor, (2) can be derived directly from the Wiener-Hopf equation.
[5] J. L. Doob, "Stochastic Processes," John Wiley and Sons, Inc., New York, N. Y., ch. 12, sec. 5; 1953.
[6] *Ibid*, ch. 11, sec. 9.

d the signal input $s(t)$ in the form given by (1) then

$$Q(\omega) = L_\alpha[e^{i\alpha\omega}]$$

ere $L_\alpha[e^{i\alpha\omega}]$ is of the form of the finite sum $\sum a_\nu e^{i\alpha_\nu\omega}$ a limit in the mean of such sums. Therefore (2) can be tten as

$$PL_\alpha\begin{bmatrix} e^{i\alpha\omega} \\ 0 \end{bmatrix} = L_\alpha P\begin{bmatrix} e^{i\alpha\omega} \\ 0 \end{bmatrix},$$

d we have to prove that L_α and P commute. Let

$$\begin{bmatrix} Y_1 \\ Y_1 \end{bmatrix} = P\begin{bmatrix} Q_1 \\ 0 \end{bmatrix}; \qquad \begin{bmatrix} Y_2 \\ Y_2 \end{bmatrix} = P\begin{bmatrix} Q_2 \\ 0 \end{bmatrix}$$

en

$$\begin{bmatrix} \alpha Y_1 + \beta Y_2 \\ \alpha Y_1 + \beta Y_2 \end{bmatrix} = P\begin{bmatrix} \alpha Q_1 + \beta Q_2 \\ 0 \end{bmatrix}.$$

erefore L_α and P commute when L_α is of the form of inite sum $\sum a_\nu e^{i\alpha_\nu\omega}$. We will prove that L_α and P mmute also when L_α is a limit in the mean of such ite sums by proving the following lemma:

If $\{x_n\}$ is a sequence of vectors in a Hilbert space H con-rging to a vector x, then the sequence $\{Px_n\}$ where is a projection from H on a subspace H_1 is also con-rgent. Moreover, if $\xi = \lim_{n\to\infty}\{Px_n\}$ then $\xi = Px$.

oof: If $x_n \to x$ (in the mean), then given any $\epsilon > 0$ ere exists an n_0 such that $\| x_n - \bar{x}_m \| < \epsilon$ for all $n, m > n_0$. For every vector y in H we have

$$\| y \| \geq \| Py \|,$$

therefore,

$$\| Px_n - Px_m \| = \| P(x_n - x_m) \| \leq \| x_n - x_m \| \leq \epsilon.$$

Because of the completeness of H_1 it follows that there exists a vector ξ in H_1 such that $Px_n \to \xi$. Therefore, for $\epsilon > 0$ and $n > n_0$ we have

$$| Px - Px_n | < \epsilon$$

and

$$\| \xi - Px_n \| < \epsilon.$$

Hence, by the triangle inequality,

$$\| Px - \xi \| < 2\epsilon$$

and $Px = \xi$.

In the proof given above we have implied that $G_s(\omega)$ and $G_n(\omega)$ be integrable $(-\infty, \infty)$. This requirement is not essential. What is important is that $G_s(\omega)$ and $G_s(\omega)|Q(\omega)|^2$ be integrable $(-\infty, \infty)$. If this requirement is met and $G_n(\omega)$ is not integrable $(-\infty, \infty)$, then we can replace the elements $ae^{i\alpha\omega}$ appearing in (4) and (5) by other elements such that the norms $\|Q\|$ and $\|Y\|$ will be finite and Q and Y will correspond respectively to linear operations on the signal and linear operations on the past of the signal plus noise.

Machine Recognition of Hand-Sent Morse Code*

BERNARD GOLD†

ummary—A transistorized special purpose digital computer ed MAUDE (Morse AUtomatic DEcoder) has been designed, lt and analyzed. This computer decodes a hand-sent Morse ssage, which is printed on a teletypewriter.

MAUDE's decisions take place at a number of different levels its "knowledge" is not only that of relative durations of dots dashes, but also of the Morse code and even of certain elemen-y properties of language.

MAUDE has successfully decoded between 90 per cent and per cent of 184 operators. A successful decoding is one which ults in a text which can be easily read by a man who knows language.

t is felt that MAUDE can be a practical piece of equipment a site with heavy traffic. Its performance will be inferior to t of a man until more sophisticated language rules, using at st a word vocabulary, are included.

* Manuscript received by the PGIT, November 11, 1958. The earch in this document was supported jointly by the Army, vy, and Air Force under contract with the Massachusetts titute of Technology.

† Lincoln Lab., M. I. T., Lexington, Mass.

I. Introduction

THE DESIGN of the Morse automatic decoder is based primarily on observations of actual hand-sent Morse code messages. It will not, therefore, be explained on other than pragmatic grounds or its history traced. Rather, the rules governing MAUDE shall be presented briefly with a subsequent interpretation of the voluminous data that was a necessary product of this project. The Appendix presents some mathematical analyses pertaining directly to MAUDE. Two companion papers are in preparation—one describes in detail the engineering design; the other is a comprehensive mathematical treatment of a statistical problem which arose from the MAUDE project.

II. Definition

Morse code is a rudimentary encoding scheme for language in which each character (letter of the alphabet,

number or punctuation sign) is represented by a unique sequence of marks and spaces. The international Morse code is given in Fig. 1.

The accepted definition of an ideal, or machine, Morse code is one for which the five elements have the following relative durations: marks—dot = 1 and dash = 3; spaces—symbol space = 1, character space = 3 and word space = 7. Symbol spaces separate marks within a character, character spaces separate characters within words, and word spaces separate words. Symbol spaces shall be referred to as short spaces, and character and word spaces as long spaces.

The three standard instruments for generating a Morse signal are 1) the simple hand key, 2) the semiautomatic key, or "bug" and 3) the automatic Morse machine. The hand key because of its simplicity and low cost is most often used. Marks (dots and dashes) are produced by depressing the key; spaces are produced by lifting the key. The durations of all marks and spaces are controlled directly by the sender. A "bug" is more difficult to control and is generally used by more experienced operators. This key has two "on" positions, one of which causes a machinelike sequence of dots to be generated. The dashes are produced manually from the other "on" position. The transmission and reception of completely automatic Morse are a routine matter which is not of concern here.

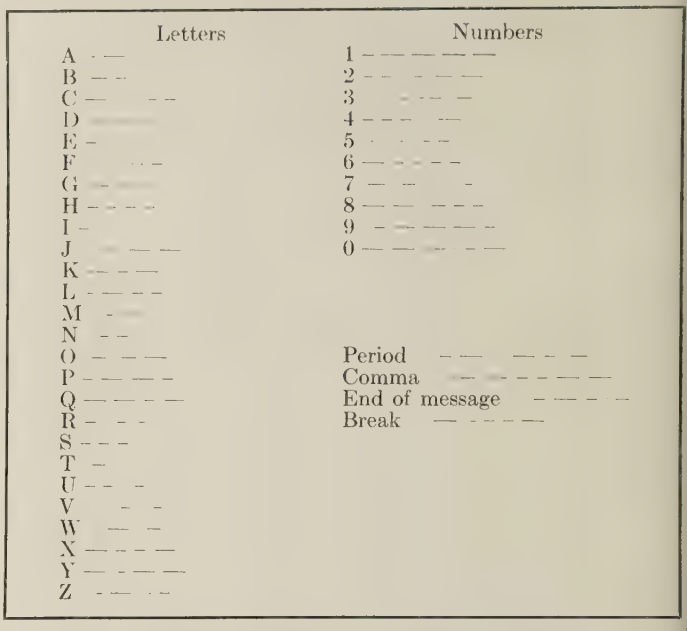


Fig. 1—The international Morse code.

III. Encoding and Language

At this point the thought implicit in the first sentence of Section II should be expanded. Morse code is itself not a language but a way of representing or coding a given language, such as English or German. It is analogous to handwriting in that there is a symbol for each character in an alphabet. In fact, some intuitive feelings about the Morse code problem may be developed by pursuing the handwriting analogy. Some people write very clearly so that anyone can read their writing and, further, even those who cannot understand the language of the clear writer may still be able to understand and reproduce each symbol (letter of the alphabet) that was written. It is, however, very helpful to know the language being written; in fact, it is usually much more difficult to read handwriting in a foreign language, even if this language is somewhat familiar, than to read one's native language in the same handwriting. Thus we see that, for handwriting, knowledge of the code is sometimes (in the case of the clear writer) sufficient to discern each intended symbol; in many cases, a higher knowledge, *i.e.*, of the language, is necessary to do so. Note that understanding of the meaning is not being discussed, but merely the identification, symbol for symbol, of what was written. This is all that is attempted in MAUDE.

A typewriter (or a Teletype machine) performs a noiseless coding; the process is exactly defined and perfectly predictable. For these codes, it is clear that only the encoding procedure need be known in order to be able to decode any message perfectly. Handwriting, speech or hand-keyed Morse code, on the other hand, are examples of noisy coding. Of these three examples, Morse code appears to be by far the easiest to decode, and yet even here it is not out of place to inquire to what extent a decoder (man or machine) needs to use only his knowledge of the code and to what extent he needs to use his knowledge of the language being encoded.

These questions will be taken up again later, after a discussion of the design and description of the operation of MAUDE.

IV. The Fixed Rules

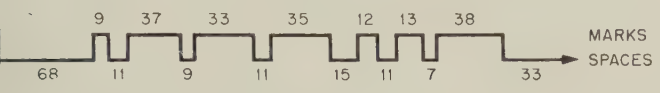
The crux of the difficulties encountered in designing MAUDE is the variability among individual senders. The main problem is to find a set of rules which will work for nearly all operators.

The fixed rules of MAUDE are such rules. Through their use, a Morse message may be partially decoded. They are rules based on the properties of the encoding scheme (*i.e.*, Morse code) and, in a rudimentary way, on the structure of language.

At this point, it was noted that a man, who knows Morse code, will never remember a sample of data as accurately as a machine; yet he decodes a Morse message well enough to make any machine thus far built turn green with envy. Why? The answer can only be that to a very great extent he knows what to expect because he knows Morse code and the language being sent.

To illustrate, the first two letters of "jumped," taken from an actual Morse code message, is considered (See Fig. 2.) The numbers refer to the relative duration of the marks and spaces. It is clear that on a purely statistical basis, the number 15 would correspond to a short space. However, the result would then be meaningless as Morse

de. In order to make it meaningful, there must be at ast one additional long space between 68 and 33. Since 5 is the longest of all other spaces, it then becomes ghly probable that 15 belongs to the set of long spaces, hich, in fact, it does.



g. 2—Morse code for J U. Top numbers are mark durations; bottom numbers are space durations. Scale: 1 unit = 5 msec.

Now the fixed encoding and language rules can be ated.

A) The longest of six successive spaces is almost always a long space. For brevity, this is called the rule L_6.

B) The shortest of six successive spaces is almost always a short space. This is the Rule S_6.

C) The shortest of three successive spaces from the set of spaces defined by L_6 is usually a character space. This is the rule S_3.

D) The longest of three, four or five successive spaces is almost always a long space if the succession of marks which they separate do not constitute a character of the Morse alphabet. (For example, four successive dashes do not constitute a single character. Thus the longest of the internal spaces is probably a long space.) This is the rule L_k, where $k = 3$, 4 or 5. This rule will also be referred to as the alphabet test.

E) The shortest of six successive marks is usually a dot. This is M_6.

All of these rules are applied to sliding intervals of arks (or spaces). In this way, the greatest number of mbols is classified.

Before a discussion of these rules, it was noted that L_6, $_k$ and S_6 can be used to categorize some of the spaces nto two sets, short and long; S_3 can help separate some of ese long spaces into character and word spaces and M_6 elps distinguish dots from dashes. A step has therefore een made towards the required ternary decision on spaces nd binary decision on marks.

The rules L_6 and L_k are based exclusively on the nature f the Morse alphabet. As seen from Fig. 1, this alphabet ontains all possible combinations of 1, 2 and 3 mark aracters, but appreciably less than half of the 4, 5 and 6 ark characters and no characters of more than 6 marks.[1] hus, at least one of any 6 successive spaces and at least ne out of certain sets of 3, 4 or 5 successive spaces is a ng space; this datum has convinced us that these long paces are very rarely of lesser duration than any of the eighboring short spaces.

The rule S_6 is based on the observation that five or more successive one-symbol characters (E or T) rarely occur in English. For random letter cipher text, the probability of occurrence is $(1/13)^5$. The data convince us that the true symbol space will not be of greater duration than nearby long spaces.

The rule S_3 is based on the rare occurrence of two successive words with no more than five internal spaces since in this event the rule L_6 may pick out three successive word spaces. The rule M_6 works well except when six successive dashes are present.

All the rules discussed are based on the structure of the encoding and the language. Another rule, based on operator sending characteristics, resulted from the discovery by Smith-Vaniz[2] that spaces following dashes tend to be shorter than spaces following dots. He suggested that a percentage of the previous marks be added to all spaces, so that new space durations $S' = S + \alpha M$, where S and M are the true durations of the space and the preceding mark, were used. In this connection, it was noted that characters ending in a dash (A, M, O, etc.) tend to be run together with the following character. In fact, the rules L_6 and L_k have, because of this, failed frequently enough to cause concern. The transformation to S' improves this situation appreciably.

The constant α, as might be expected, is really a function of the operator. However, $\alpha = 0.18$ has been found to result in a fairly universal improvement for hand-sent Morse. The transformation is more of a hindrance than a help for "bug" sending, yet not harmful enough to deterioriate noticeably the performance of MAUDE.

It is clear from this discussion that these rules are not absolute. However, their application has proved very helpful.

V. The Threshold Tests

Three thresholds, T_M, T_S and T_C are used (see Fig. 3) to classify the remaining marks and spaces (those not picked by the fixed rules). These thresholds will fluctuate during a given message as well as from message to message and for different operators. θ_M, θ_S and θ_C are running averages of those marks and spaces "tagged" by the language rules and thus the θ's should accurately mirror the properties (such as word speed) of the particular message. Δ_M, Δ_S and Δ_C are empirically determined constants,[3] whose values, ideally should be such that the T's are always at their optimum setting.

The situation is described statistically in Fig. 4. Let f_1 and g_1 be the probability density functions for the short and long spaces. Decisions are now made by the rules S_6 and L_6 and those spaces remaining are shown by

[1] If the error sign is called a character, it is the one exception to this statement. Section VI describes the special treatment needed for error signs.

[2] Private communication.

[3] In Fig. 2, all variables are actually logarithms of time durations. This is accomplished in the equipment by transforming the durations of all marks and spaces so that $x = \log_{16/15} y$ where y is the original duration. The Δ's in Fig. 2 thus correspond to ratios, which, it is felt, contain less inherent fluctuation than differences. Also, the dynamic range is increased.

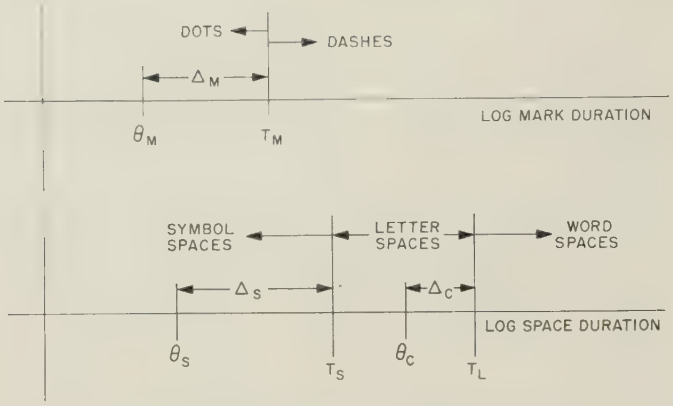


Fig. 3—Thresholds used in mark and space decision.

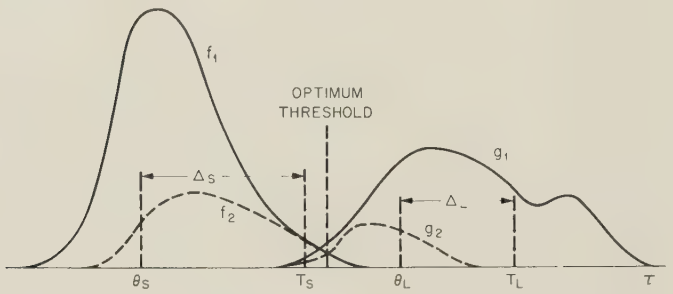


Fig. 4—Density functions f_1 and g_1 for all short and long spaces, respectively. f_2 and g_2 indicate the distribution of spaces remaining unidentified after the operation of rules L_6 and S_6.

f_2 and g_2. A threshold T_S which adequately separates f_2 and g_2 must now be defined.

Empirically, the following rule has been established: Let $\theta_S(n)$ be defined by,

$$\theta_S(n) = \theta_S(n-1) + \frac{1}{\beta}\left[x_S(n) - \theta_S(n-1)\right] \qquad (1)$$

with the initial condition $\theta_S(1) = x_S(1)$. $x_S(n)$ is the nth space picked by S_6. $\theta_S(n)$ can thus be considered as a weighted average of the first n spaces "tagged" by S_6. In fact, (1) can be solved by iteration to give

$$\theta_S(n) = x_S(1)\xi^{n-1} + (1-\xi)\sum_{2}^{n} x_S(i)\xi^{n-i} \qquad (2)$$

where $\xi = 1 - 1/\beta$. The value $\beta = 4$ has been used.

A simple calculation shows that $\bar{\theta}$, the mean of $\theta_S(n)$, is always the same as $\bar{x}$, the mean of all the $x_S(i)$. Thus $\theta_S(n)$ is the sample mean of the density function defined by f_2.

As ascertained from data,

1) T_S can be formed by adding a fixed value of Δ_S to θ_S, with generally good results.
2) If Δ_S is chosen for each operator and then fixed for the entire message, a worthwhile improvement over 1) is obtained. A detailed analysis of data based on both 1) and 2) is given in Section IX.

Similar arguments hold for the thresholds T_C and T_M. Since the rules M_6 and S_3 are not as reliable as S_6 and L_6, the former are used only to generate θ_M and θ_L. The thresholds T_M and T_C are then used to make all decisions.

VI. The Error Sign

When an operator makes an error and realizes it, he sends an error sign. This most commonly consists of a sequence of 6 to 20 dots. However, practices vary—sometimes the characters IMI are sent, other times simply II is sent.

If the fixed rules were applied to the common error sign, the values of the thresholds would be seriously affected. Since error signs occur fairly often, it was necessary to have an error sign detector whose output (when an error sign occurred) could inhibit the remainder of our system.

Upon observation of the properties of actual error signs, the following criterion was empirically established: if five or more successive logarithmic differences between adjacent marks and between adjacent spaces are each less than a fixed number, an error sign is present. Recent experiments indicate that a fixed logarithmic difference of 10 (corresponding to a ratio 2:1) does the job.

VII. Sequence of Operations

The laboratory model of MAUDE is a special purpose digital computer using transistorized flip-flops as the basic storage unit. A block diagram is shown in Fig. 5. In this section the sequence of operations shall be described briefly. A fuller description is contained in a companion paper.

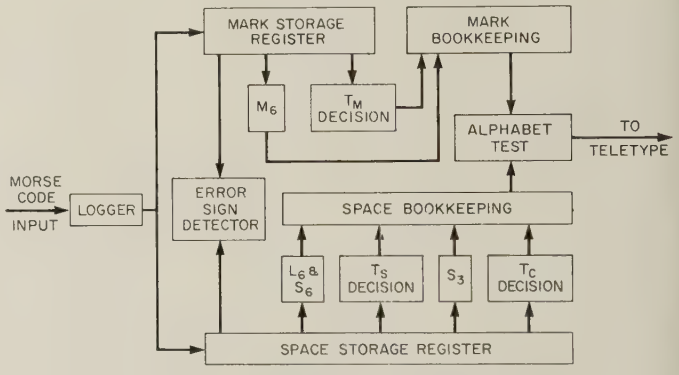


Fig. 5—Block diagram of MAUDE.

The logarithms of the durations of alternate marks and spaces are stored digitally in the logger. The counts representing these logarithms are alternately passed through the mark and space storage registers.

Rules L_6, S_6 and M_6 are first applied. The marks picked by M_6 are used to adjust a θ_M counter as indicated by (2). In the present model, all mark decisions are made by comparing each mark with $\theta_M + \Delta_M$, the latter being fixed. Decisions between short and long spaces are made by applying L_6 and S_6 followed by a threshold test. S_3 is performed in supplementary storage.

Fig. 6 shows in detail how parts of two actual messages are decoded by MAUDE. The numbers on the marks line are the logs of the actual durations of marks, while the space numbers are the results of the transformation

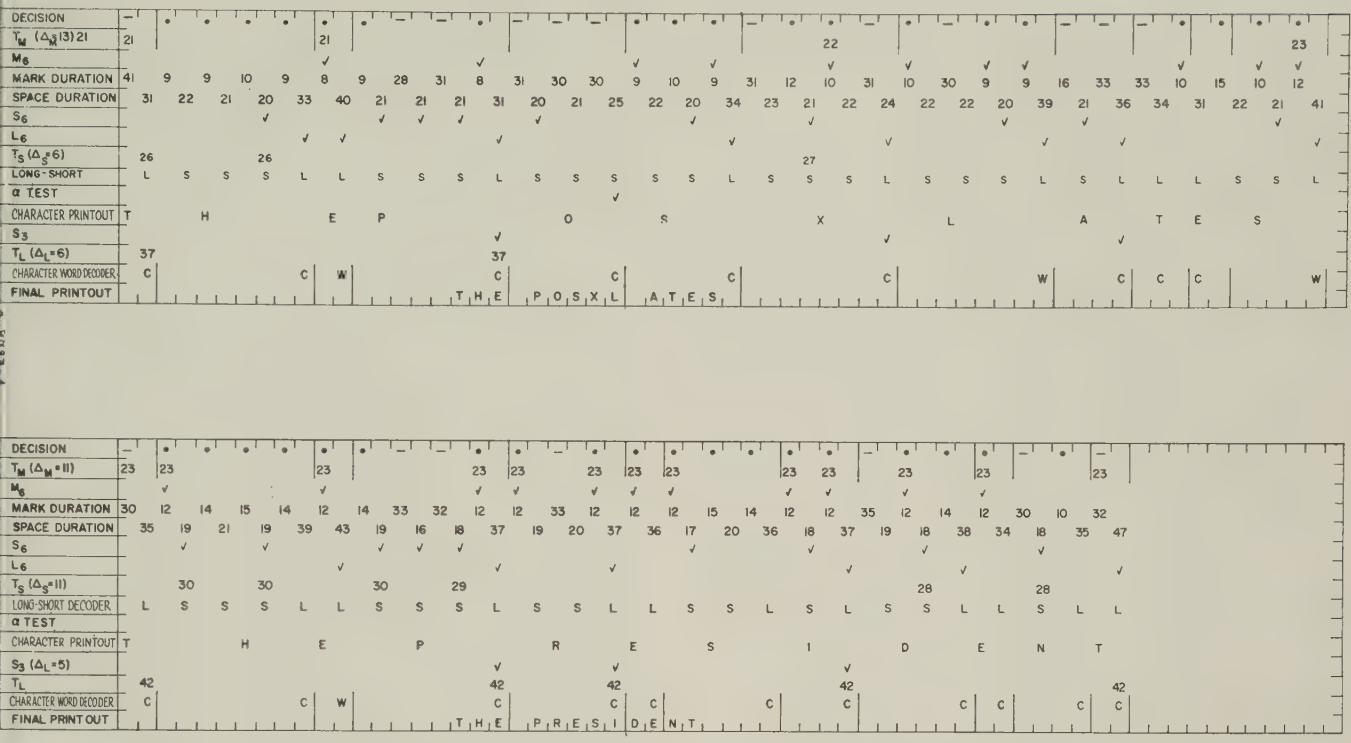


Fig. 6—Detailed operation of MAUDE on actual message.

$S' = S + kM$, described in Section V. By following the check marks, the reader can see which marks and spaces were picked by the various rules and can mentally compute the thresholds.

The alphabet test successfully separated O from S in postulates because — — — ··· does not exist in our Morse alphabet. However, it failed to separate T from U because X (TU run together) is in the Morse alphabet. This mistake could not be rectified on a letter basis, but would require more complicated contextual rules.

It should be noted that the man who sent the first message tends to run letters together more often than most operators, yet he is by no means extraordinary, and an experienced receiving operator would not have much trouble decoding this message. It might also be mentioned that other operators do the opposite, *i.e.*, tend to split characters.

The second operator was decoded perfectly by MAUDE. This is fairly typical for about half of the operators that were decoded by MAUDE.

VIII. Classification of Errors

The types of errors that either a man or a machine would make in decoding a Morse message depend on the characteristics peculiar to the sender. A classification of the errors encountered in MAUDE translations will now be given in the hope that it will give a fairly inclusive picture of the errors to be expected.

1) Failure of the rule L_6. Examples: $OF \rightarrow TGR$, $OF \rightarrow 8N$, $JU \rightarrow EMNA$, $OP \rightarrow TGG$. In these cases, the sender paused for less time after the third dash in O (or J) than after other marks in the sequence. Thus L_6 picked the wrong space. L_6 did not fail often enough to make revision of the rule necessary. It was noted, however, that most of the failures involved the letter "*O*."

2) Failure of the alphabet test. Examples: $OR \rightarrow MC$, $OU \rightarrow MX$, $OS \rightarrow MB$. The cause of failure is exactly the same as for L_6. The alphabet test was applied, as it should have been, but the wrong space was the longest.

3) Two characters run together. Examples: $TH \rightarrow 6$, $MA \rightarrow Q$, $AI \rightarrow L$, $AN \rightarrow P$, $ON \rightarrow 9$, $TW \rightarrow Y$. In these cases, the letter space, *e.g.*, between T and H, was too short. A decision could be made only on a threshold basis since the alphabet test could not yield a unique decision.

4) Split characters. Examples: $A \rightarrow ET$, $U \rightarrow EA$, $H \rightarrow II$, $J \rightarrow EO$, $O \rightarrow MT$, $L \rightarrow AI$. This is the inverse of 3). The sender pauses too long during a character. Again, only a threshold decision basis is possible.

5) Simple mark errors. Examples: $A \rightarrow M$, $F \rightarrow H$, $J \rightarrow P$, $G \rightarrow O$, $O \rightarrow G$. These are caused by incorrect threshold decisions on a dot or dash in a character.

6) Complex mark errors. Examples: $F \rightarrow UT$, $P \rightarrow MN$, $Q \rightarrow MM$, $Y \rightarrow OT$. In these cases, a mark error (based on the threshold) produces an impossible sequence of dots and dashes. This automatically sets the alphabet test to work. In the present model of MAUDE the alphabet test can only change short spaces to long spaces. Hence, a mark error generates a space error.

It is important to note that all of the above errors were caused by fluctuations in the durations of marks or spaces. Thus, it is reasonable to expect that many of the wrong decisions which caused these errors were close decisions, *e.g.*, a slight change in the threshold value may have

reversed a decision. For the data analyzed, about half of the total errors were of the types 1) through 6).

7) Extra dot (simple case). Examples: $D \rightarrow B$, $E \rightarrow I$, $M \rightarrow W$, $I \rightarrow S$, $O \rightarrow Q$, $S \rightarrow H$, $U \rightarrow V$. The sender occasionally sends an extra dot, through carelessness or fatigue. This happens more often with "bug"-sent messages. It can also result from "key bounce" on hand sent or bug-sent Morse.

8) Extra dot (complex cases). Extra dots most often occur when numbers are sent. This triggers the alphabet test so that, *e.g.*, $6 \rightarrow DI$.

9) Missing dot (simple). Examples: $V \rightarrow U$, $B \rightarrow D$, $D \rightarrow N$, $F \rightarrow R$, $X \rightarrow K$, $J \rightarrow O$, $L \rightarrow R$.

10) Missing dot (complex). Examples: $C \rightarrow TN$, $D \rightarrow TE$, $Q \rightarrow MT$. In these cases a longer space is generated where the dot was left out.

11) Extra space. Examples: $C \rightarrow 6$, $Q \rightarrow 7$. A dash was broken up into two dots by the insertion of a space during the dash.

12) Missing space. Examples: $A \rightarrow T$, $B \rightarrow K$, $O \rightarrow M$, $H \rightarrow U$, $U \rightarrow M$. Here, either two dots are run together to make a dash, or a dot-dash creates a single dash.

13) Complex errors. About 10 per cent of the garbles analyzed could not be attributed to any simple errors, *e.g.*, $THE \rightarrow XT$, $TH \rightarrow KB$, $AC \rightarrow Y$, $JUMP \rightarrow WXTY$.

Errors of the type 7) through 13) can be caused by operator sending lapses and also by noisy radio reception, *e.g.*, fading, atmospheric noise, adjacent channel interference, etc. It is more difficult to correct such errors because the origin of the garble cannot be easily located.

IX. Analysis of MAUDE Performance

Since operator variability causes the main difficulty, any judgment on a Morse decoder must be based on data from many different operators. Furthermore, there must be some idea as to how the error rate affects the ability of the human observer to understand the messages. For example, if MAUDE decoded all messages with no more than 2 per cent of the characters garbled, there would not be too much point in improving MAUDE to make the maximum error rate 1 per cent since it has been ascertained that people can correct texts containing 2 per cent and 1 per cent error with almost equal ease.[4]

It is felt that the effectiveness of a machine such as MAUDE depends on the percentage of received messages that it can adequately decode. The word "adequately" pinpoints the vagueness of this statement; this is interpreted to mean "capable of reconstruction in a reasonable length of time;" even here the problem becomes too subjective. Miller's paper and the tests run by Van Hoosen at Lincoln Laboratory[5] give some insight into this problem and the interested reader may refer to these works.

[4] Referred to, of course, is texts of a real language that the observer knows. See G. Miller and E. Friedman, "The reconstruction of mutilated English texts," *Information and Control*, vol. 1, September, 1957.

[5] Van Hoosen, Lincoln Lab., M. I. T. Lexington, Mass., Quart. Prog. Rep.; July 15, 1958. (Not generally available.)

Table I shows the total number of messages with less than a given percentage of character errors out of 184 messages sent by 53 different operators (containing about 45,000 characters). Table II shows a similar compilation for space errors, defined as a letter space mistaken for a word space or vice versa. Table III shows error rates for 12 operators for whom English text, number cipher text and letter cipher text were obtained. There was no significant difference in the quality of the different types of texts as translated by MAUDE.

TABLE I
Number of Messages, N, Having Less Than the Given Per Cent Character Errors

Per Cent Character Errors	N
0.25	43
0.5	58
0.75	76
1	87
1.5	111
2	127
3	139
4	145
5	149
6	154
7	157
8	164
9	166
10	169
over 10	184

TABLE II
Number of Messages, N, Having Less Than the Given Per Cent Space Errors

Per Cent Space Errors	N
1	5
2	18
3	34
4	53
5	71
10	109
15	123
20	124

TABLE III
Character Error Rates, by Operator, for Three Types of Message. The Error Rate is Given in the Form A/B, Where A is the Number of Character Errors (All Types) and B the Total Number of Characters Sent.

Operator No.	English	Letter Cipher	Number Cipher
1	4/655	7/1068	9/687
2	5/459	4/678	1/475
3	8/655	8/678	19/687
4	8/459	29/913	10/475
5	7/655	2/865	3/687
6	20/655	10/703	5/475
7	7/655	6/1068	7/687
8	2/655	5/703	0/687
9	5/449	12/858	10/687
10	1/253	7/1108	1/687
11	1/449	6/1108	3/502
12	4/459	10/1023	3/352
Composite	72/6438 = 1.12 per cent	106/10772 = 0.99 per cent	71/7088 = 1 per cent

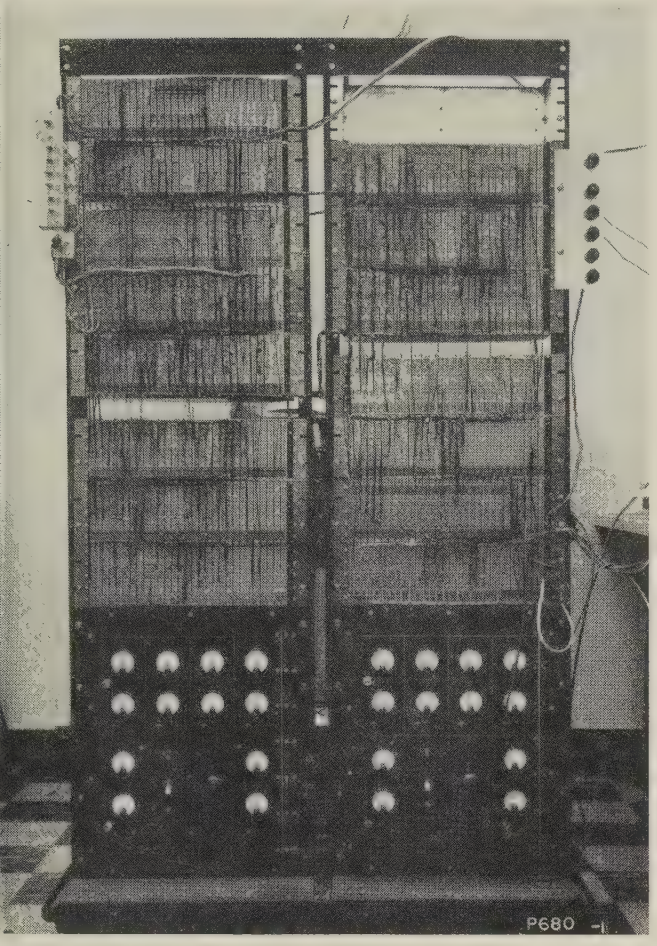


Fig. 7—Laboratory model of MAUDE.

For purposes of discussion, assume that a message ving less than 6 per cent character errors is acceptable.[6] n this basis, note from Table I that 30 out of 184 messages e not acceptable. Now, Table I was based on a fixed Δ sign for MAUDE, in which Δ_M, Δ_S and Δ_C were chosen r analyzing about 10 of the messages and picking the best lues for them. All 184 messages were then run off keep- g all the Δ's fixed.

It is clear from Table I that a rather high percentage of e total messages was badly garbled. In order to investi- te this more carefully, a simulation program on the IBM 4 was run first in which the MAUDE program was neralized so that it could choose its own value of Δ for ch 100 samples of marks and spaces. In order to make a rect comparison between the fixed and variable Δ hemes, errors only of the types 1) through 6) were unted and, on that basis, another table of cumulative stributions was constructed. (The errors 7) through 12) e completely unaffected by the type of variation on Δ, so ey were not included in the comparison. They are listed the third column of Table IV and will be discussed in a oment.) From the first two columns of Table IV, a ticeable difference between the fixed and variable Δ

[6] If willing to use such a crude criterion, then 6 per cent is the st guess based on Van Hoosen's data. The combined effect of aracter and space errors has not yet been checked, but the latter ms appreciably less significant.

TABLE IV
Number of Messages, N, Having Less Than the Given Per Cent Character Errors

Errors 1) through 6)			Errors 7) through 12)
Per cent	Fixed Δ	Variable Δ	Incorrect Number of Marks
0.25	73	84	72
0.5	87	104	79
0.75	110	129	107
1	120	138	122
1.5	132	147	137
2	145	154	150
3	158	164	162
4	162	170	171
5	164	175	175
6	164	177	178
7	167	179	180
8	169	180	181
9	169	181	181
10	170	181	181
Over 10	184	184	184

schemes can be seen. For example, using the 6 per cent criterion, twenty messages are unacceptable for fixed Δ's, only seven for variable Δ's.

From column 3 of Table IV, note that messages in which only the type of errors 7) through 12) were counted caused about the same number of unacceptable messages as was caused by the variable Δ program. No way is known to correct these errors without the use of more complex contextual rules involving knowledge of word structure. Furthermore, even if all the errors of column 2 were eliminated and only those of column 3 remained, there would still be a significant difference in performance between a man and MAUDE. The conclusion is inescapable, therefore, that for the automatic reception of a language encoded by even a simple process like Morse code, a machine must have some knowledge of the language if it is to approximate the performance of a man. In particular, MAUDE must, to some extent, know English if it is to decode English sent by Morse code.

The reader will note a return to the questions raised in Section III. If allowed to generalize somewhat from the experience with MAUDE, it would be concluded that efforts to decode by machine "noisily" encoded language symbol by symbol, *e.g.*, speech recognizers using only the waveform of the speech, are bound to be severely limited. It must be recognized that better decoding machines depend not only on ingenious decoding schemes, but also on more use of the structure of the language and greater understanding of human decoding processes.

Appendix

A fairly complete mathematical theory of the rules L_6, S_6 and M_6 has been worked out and a separate paper has been submitted for publication. This Appendix is concerned with the part of the theory applied most directly to MAUDE which, in turn, is based primarily on a generalization of a theorem due to Maximon.[7]

[7] L. C. Maximon, "Some Properties of Infinite Sequences of Independent Random Variables," Lincoln Lab., M.I.T., Lexington, Mass., Group Rep. 34–53, 1956. (Not generally available.)

Theorem: Let y_m be the mth term ($m \geq k$) in any sequence Y of independent random variables. Let Y be operated on by the maximum rule of k (an integer ≥ 2) to produce a subsequence V. The probability $P(y_m)$ that y_m be in V is

$$P(y_m) = \sum_{i=1}^{k} p(R_i) - \sum_{i=1}^{k-1} p(R_iR_{i+1}) \tag{3}$$

where R_i is the set of numbers $y_i, y_{i+1}, \cdots, y_{i+k-1}$, R_iR_{i+1} is the set of numbers $y_i, y_{i+1} \cdots, y_{i+k}$, $p(R_i)$ is the probability that y_m is the largest number in R_i, $p(R_iR_{i+1})$ is the probability that y_m is the largest number in R_iR_{i+1}. Note that it is identical for both minimum and maximum rules.[8]

To apply the theorem to a Morse signal, probabilities q and $1 - q$ are assigned to the occurrence of short and long spaces (or dots and dashes). It can be assumed that a long space is always longer than a short space in a given set R_i. Then, the probability $p(R_i)$ that a given long space is the longest long space in the set R_i can be shown to be

$$p(R_i) = \sum_{j=1}^{k} \frac{q^{k-j}(1-q)^{j-1}}{j} \binom{k-1}{j-1} = w \tag{4}$$

and

$$p(R_iR_{i+1}) = \sum_{j=1}^{k+1} \frac{q^{k+1-j}(1-q)^{j-1}}{j} \binom{k}{j-1} = z. \tag{5}$$

Using (3) and noting that $p(R_i)$ is invariant with i,

$$P(y_m) = kw - (k-1)z. \tag{6}$$

Eq. (6) is tabulated in Table V.

TABLE V
TABULATION OF (6).

k	$q = \frac{1}{2}$	$q = \frac{2}{3}$	$q = \frac{3}{4}$
2	0.91667	0.9630	0.9792
3	0.8125	0.9074	0.9453
4	0.7125	0.8444	0.9039
5	0.6250	0.7888	0.8594
6	0.5513	0.7151	0.8113

ACKNOWLEDGMENT

Many people contributed to the MAUDE project. The original impetus came from O. G. Selfridge. R. Berg, who will submit a paper on the design of MAUDE, was chiefly responsible for the engineering work. B. Eisenstadt, P. Fleck, W. McLaughlin and D. Nelson contributed to major portions of the design. C. McElwain and B. Byrnes simulated MAUDE on the IBM 704. T. S. Pitcher made many valuable suggestions and M. Freimer, L. Maximon and A. Tritter contributed to the theoretical analysis of the rules L_6 and S_6. M. Balser very kindly edited the text and improved its style and comprehensibility.

[8] This theorem is proved in Eisenstadt, Gold, *et al.*, "MAUDE (Morse Automatic Decoder)," Lincoln Lab., M.I.T., Lexington, Mass., Group Rep. 34–57; 1957.

The Morse Distribution*

M. FREIMER†, B. GOLD† AND A. L. TRITTER†

ummary—**A problem which arose during research involved in igning a machine to translate hand-keyed Morse code into ted text may be stated as follows: Let $X = \{x_i : i = 1, 2, \cdots, n\}$ a sequence of independent random variables all of which have same distribution. Assume that the probability that $x_i = x_j$, $\neq j$, is zero. Let k be a positive integer $\leq n$, and consider all sequences $x_i, x_{i+1}, \cdots, x_{i+k-1}$ of X consisting of k consecutive iables. Let us distinguish, with a check ($\surd$), the largest member ach such subsequence. We have studied, and partially tabulated, (r), the probability that exactly r members of the sequence X *not* checked. This paper contains most of the pertinent results.**

Introduction

MORSE code (as sent by hand) is one of the simplest of aural languages, yet it has features common to all spoken languages. At the sug-stion of O. G. Selfridge, we have studied ways of building machine which can recognize Morse code. We have mulated certain rules, both linguistic and statistical, ich have proved very valuable for our purpose. Analysis these rules led us to investigate a discrete probability tribution, apparently never before studied, which we all call the Morse distribution. In this paper we present tain mathematical properties of the Morse distribution, ich is tabulated in the Appendix.

Linguistic Rules

Morse code is a sequence of alternating marks and aces. Marks are either dots or dashes. For our purposes aces are categorized into two groups: a) short spaces, ich separate two marks of the same character—letter, mber or punctuation mark—and b) long spaces, of ger duration, which separate adjacent characters.

No character in the Morse alphabet contains more than marks (and thus, 5 short spaces). We would therefore pect, and experience has proven, the following rule to true:

le 1—The largest of any group of 6 successive spaces (almost surely) a long space.

The second rule follows from an elementary property the English language (shared, it seems, by others), *i.e.*, at a sequence of 5 successive letters, each of which is or T, is very improbable.[1] The Morse Code for E is a gle dot; for T, a single dash; any other character must ntain at least two marks and therefore at least one short space. We thus establish the second rule:

Rule 2—The smallest of any group of 6 successive spaces is (almost surely) a short space.

One of the primary functions of our machine is to select the largest and smallest of each sequence of 6 successive spaces and categorize those selected according to the above rules. One of the questions for which we seek an answer is, "What are the statistical properties of the number of spaces selected by these rules?" We have tried to answer this question both experimentally and theoretically; in this paper we present the results of the theoretical investigation.

The Morse Distribution

We define a discrete probability distribution, which we have called the Morse distribution, as follows.

Let $X = \{x_i: i = 1, 2, \cdots, n\}$ be a sequence of independent random variables, all having the same continuous probability distribution $\lambda(x)$. For any positive integer $k \leq n$, define a new sequence $Z = \{z_i: i = 1, 2, \cdots, n\} = L_k(X)$ such that $z_i = 1$ if x_i is the largest[2] of any set of k consecutive variables of X, and $z_i = 0$ otherwise.[3] An example, for $k = 3$, is shown in Fig. 1. In this example x_2 is the largest member of the sets S_1 and S_2; and x_5 is the largest member of the set S_3.

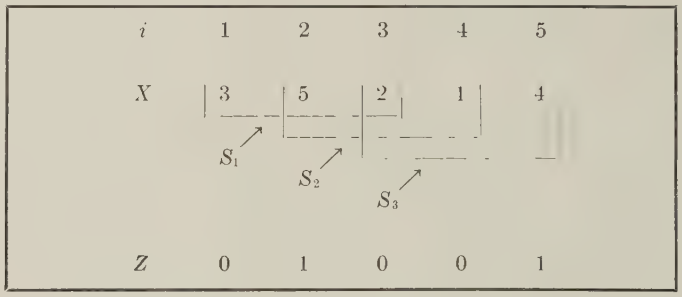


Fig. 1—Example of the transformation $Z = L_k(X)$.

Letting $r = n - \sum_{i=1}^{n} z_i$, we define the probability of r, for fixed n and k, as $p_n^k(r)$. This is the general term of the Morse distribution. Physically, r represents the number of variables (in X) *not* picked out of a sequence of length n, by the operation L_k defined above or, in other words, the number of zeros in Z.

* Manuscript received by the PGIT, January 26, 1959.

† Lincoln Lab., M.I.T., Lexington, Mass. The work reported e was performed at the Linclon Laboratory, Massachusetts titute of Technology, with the joint support of the Army, Navy d Air Force, under contract.

[1] But not impossible, *e.g.*, "sweet teeth."

[2] The problem is the same whether the largest or smallest of k consecutive variables is to be chosen. For convenience, therefore, we shall always consider only the largest.

[3] For $k = 1$ this is a trivial procedure, since each $z_i = 1$. Hence we shall only consider $k \geq 2$.

Since we have assumed $\lambda(x)$ to be continuous, the probability of $x_i = x_j$ for any pair $i \neq j$ is zero. If we ignore this null event then in any sequence X the variables x_i can be ordered (by magnitude) in a unique way. We may therefore replace X by a sequence Y, the terms of which are the integers 1 through n in the naturally corresponding order. If we apply L_k to Y we get the same sequence Z.

We can now compute

$$p_n^k(r) = \frac{f_n^k(r)}{n!} \tag{1}$$

where $f_n^k(r)$ is the number of permutations of the n integers which are carried by L_k into those sequences, Z containing exactly r zeros. We thus see that the distribution function $\lambda(x)$ plays no part in computing the probabilities $p_n^k(r)$.

We shall omit the superscript k when this causes no confusion.

Recursion Formulas for $f_n(r)$ and $p_n(r)$[4]

Let us suppose that the largest number, *i.e.*, n, is the $i + 1$ term in Y. Then, the part of Z formed from the i numbers to the left of n is *independent* of the part formed by the $n - i - 1$ numbers to the right of n. Furthermore, these first i places of Z are the same as those that would be obtained by applying L_k to the sequence of total length i that is identical with the first i places of Y, and similarly for the last $n - i - 1$ places of Z. Finally, there are $\binom{n-1}{i}$ ways of choosing the numbers in the first i places from $1, 2, \cdots, n-1$. From these considerations, we derive the following recursion formulas:

$$\left.\begin{aligned} f_n(r) &= \sum_{i=0}^{n-1} \sum_{j=0}^{r} \binom{n-1}{i} f_i(j) f_{n-i-1}(r-j), \quad n \geq k \text{ (a)} \\ p_n(r) &= \frac{1}{n} \sum_{i=0}^{n-1} \sum_{j=0}^{r} p_i(j) p_{n-i-1}(r-j), \qquad n \geq k \text{ (b)} \end{aligned}\right\}. \tag{2}$$

For $n < k$, we adopt the convention that nothing is picked; hence, all terms of Z are zero and

$$p_n(r) = \delta_{nr} = \begin{cases} 0, & n \neq r \\ 1, & n = r. \end{cases} \tag{3}$$

This convention is necessary in order that (2) be correct.

Eqs. (2) make it possible to compute $p_n(r)$ numerically for successively larger values of n. A much simpler form than (2) can be obtained by defining generating functions

$$a_n(y) = \sum_{r=0}^{\infty} p_n(r) y^r = \sum_{r=0}^{n} p_n(r) y^r, \tag{4}$$

which satisfy the equations

$$a_n(y) = \frac{1}{n} \sum_{i=0}^{n-1} a_i(y) a_{n-i-1}(y), \qquad n \geq k. \tag{5}$$

To prove (5), we multiply both sides of 2(b) by y^r and sum over r from 0 to ∞, then interchange the order of the r and j summations on the right-hand side.

From (3) and (4), we obtain

$$a_n(y) = y^n \qquad n < k. \tag{6}$$

In the Appendix, we have tabulated $p_n^k(r)$, obtained from (4) and (5), for $k = 2, 3, 4, 5, 6$, and $n \leq 13$. (For larger n, computation becomes increasingly laborious.) In addition, we have derived the following explicit expressions for $a_k(y)$, $a_{k+1}(y)$, and $a_{k+2}(y)$ for all k:

$$\left.\begin{aligned} a_k(y) &= y^{k-1} && \text{(a)} \\ a_{k+1}(y) &= \frac{2y^{k-1} + (k-1)y^k}{k+1} && \text{(b)} \\ a_{k+2}(y) &= \frac{4y^{k-1} + 4ky^k + (k+1)(k-2)y^{k+1}}{(k+1)(k+2)} && \text{(c)} \end{aligned}\right\}. \tag{7}$$

Differential Equation for the Generating Function $F(x, y)$

We can derive a simple recursion formula for $a_n^2(y)$ by defining the generating function

$$F(x, y) = \sum_{n=0}^{\infty} \sum_{r=0}^{\infty} p_n(r) x^n y^r = \sum_{n=0}^{\infty} a_n(y) x^n. \tag{8}$$

If we multiply both sides of 2(b) by $x^n y^r$ and perform a series of manipulations, we derive the differential equation

$$\frac{dF}{dx} + (1 - y) \sum_{n=0}^{k-2} (n+1) x^n y^n = [F(x, y)]^2 \tag{9}$$

with boundary condition $F(0, y) = 1$ for all y.

For $k = 2$, the solution of (9) is

$$\begin{aligned} F(x, y) &= \frac{1 + e^{2xw} z}{1 - e^{2xw} z} \\ &= w[1 + 2(e^{2xw} z + e^{4xw} z^2 + e^{6xw} z^3 + \cdots)] \end{aligned} \tag{10}$$

where

$$w = \sqrt{1 - y}, \qquad z = \frac{1 - \sqrt{1 - y}}{1 + \sqrt{1 - y}} = \frac{1 - w}{1 + w}.$$

From (8), we note that

$$a_n(y) = \frac{1}{n!} \frac{\partial^n F}{\partial x^n} \bigg|_{x=0}.$$

[4] The recursion formulas (2) and the differential equation (9) were first derived by T. Austin, R. Fagen, T. Lehrer, and W. Penney, "The distribution of the number of locally maximal elements in a random sample from a continuous distribution," *Ann. Math.* vol. 28, secs. 1–3, pp. 786–790; 1957. Their work, motivated by the formulation of the problem, contributed to the development of this analysis.

manipulating (10) we arrive at the formula (valid $k = 2$)

$$a_{n+1}(y) = \frac{2y(1-y)}{n+1}\frac{d[a_n(y)]}{dy} + ya_n(y). \qquad (11)$$

(11) lends itself readily to computation of $p_n^2(r)$.

For $k = 3$, the solution to the differential equation (9) volves Bessel functions (of $\frac{1}{3}$ integral orders). For $= 4$, Weber functions are obtained; for $k \geq 5$, the nctions are unnamed (and untabulated). In none of ese cases could we derive an expression analogous to (11).

Asymptotic Property of $p_n^k(r)$

If all x_i in X are independent, then in Z, z_i is independent z_{i+m}, provided $|m| \geq 2k - 1$. The Morse distribution (r) then tends to the Gaussian[5] for large n, and we ould expect that this happens sooner (with respect to e size of n) for smaller k.

It remains, therefore, to find the first two moments of (r). With these, and (5) and (6), we will be in a position obtain a fairly accurate approximation to $p_n^k(r)$.

Moments of $p_n^k(r)$

We shall prove that the mean and dispersion of r are ven by

$$(r)\begin{cases} = n & 0 \leq n \leq k-1 \text{ (a)} \\ = \dfrac{k-1}{k+1}(n+1) & n \geq k \text{ (b)} \end{cases} \qquad (12)$$

$$= (n+1)\left[\frac{8}{k+1}\sum_{j=k}^{2k-1}\frac{1}{j+2} - \frac{2k(5k+3)}{(2k+1)(k+1)^2}\right] \qquad n \geq 2k \qquad (13)$$

$$\sigma_n^2 = C_k(n+1) \qquad n \geq 2k. \qquad (13')$$

To prove (12), we note first that $E_n(r) = \sum_{r=0}^{\infty} rp_n(r)$. rom 2(b) we can, with some manipulation, show that

$$E_n(r) = \frac{2}{n}\sum_{i=0}^{n-1} E_i(r) \qquad n \geq k. \qquad (14)$$

We use (3) to establish 12(a), and prove 12(b) in- uctively, using (14).

To prove (13), we note that $\sigma_n^2 = E_n(r^2) - (E_n(r))^2$, here $E_n(r^2) = \sum_{r=0}^{\infty} r^2 p_n(r)$ is the expected value of r^2. gain, from 2(b) we derive

$$= \frac{2}{n}\sum_{i=0}^{n-1}\sigma_i^2 + \frac{1}{n}\sum_{i=0}^{n-1}[E_i(r) + E_{n-i-1}(r)]^2 - (E_n(r))^2$$

$$n \geq k. \qquad (15)$$

[5] S. Bernstein, "Sur l'extension du théoreme limité de calcul des obabilités aux sommes des quantités dépendentes," *Ann. Math.*, l. 97, pp. 1–59; 1927.

If $n \geq 2k$, we can evaluate the second sum on tue right-hand side of (15). It equals

$$\frac{2k(k-1)}{3(k+1)} + n(E_n(r))^2,$$

and (15) becomes

$$\sigma_n^2 = \frac{2}{n}\left[\sum_{i=0}^{n-1}\sigma_i^2 + \frac{k(k-1)}{3(k+1)}\right] \qquad n \geq 2k. \qquad (16)$$

If we multiply (16) by n, add $2\sigma_n^2$ and divide by $n+1$, it becomes clear that

$$\frac{\sigma_n^2}{n+1} = \frac{\sigma_{n-1}^2}{n+2} = C_k \qquad n \geq 2k \qquad (17)$$

where C_k depends only on k. We can now, for convenience, evaluate C_k by evaluating σ_{2k}^2. We do this by going back to (15) and, after some lengthy manipulation, obtain σ_{2k}^2, hence (13′). We have computed

$$C_2 = \frac{2}{45}, \quad C_3 = \frac{23}{420}, \quad C_4 = \frac{29}{525}, \quad C_5 = \frac{1097}{20790},$$

and

$$C_6 = \frac{15611}{315315}.$$

By using the well-known function[6]

$$\psi(z) = \frac{\Gamma'(z)}{\Gamma(z)}$$

we can perform the sum appearing in (13), *i.e.*,

$$\sum_{j=k}^{2k-1}\frac{1}{j+2} = \psi(2k+2) - \psi(k+2).$$

It is known that

$$\psi(z) = \log(z) + O\left(\frac{1}{z}\right)$$

for large z, and we thus obtain

$$C_k = \frac{8\log 2 - 5}{k} + O\left(\frac{1}{k^2}\right), \qquad k \to \infty. \qquad (18)$$

The End Point Problem

The Z sequence is stationary between the limits $k \leq i \leq n-k$, but for $i < k$ or $i > n-k$ even the individual probabilities that $z_i = 0$ vary with i. A completely stationary sequence can be derived by omitting the end points. First, define a sequence $U = (u_i)$, $i = 2-k, \cdots, n+k-1$, and a corresponding sequence $V =$

[6] See "Higher Transcendental Functions," vol. 1, sec. 1.7, formulas 1 and 10, and sec. 1.18, formula 7, Bateman Manuscript Project, McGraw-Hill Book Co., Inc., New York, N. Y.; 1953.

(v_i), $i = 2 - k, \cdots, n + k - 1$, derived from U in the same way that Z was derived from X. The subsequence $V' = (v_i)$, $i = 1, \cdots, n$ will then be stationary. Let $q_n^k(r)$ be the probability that V' has r zeros.

We have succeeded in showing that

$$p_{n+2}^2(r+1) = q_n^2(r) \qquad n \geq 0. \tag{19}$$

For $k \geq 3$, we have found no comparable equivalence. For fixed k, it is clear that in the limit of large n, the end points assume less significance and $p_n(r)$ and $q_n(r)$ approach each other.

Mean Value of r for V'

Defining the mean of r in V', $E_n'(r) = \sum_{r=0}^{\infty} r q_n(r)$, we shall show that

$$E_n'(r) = \frac{k-1}{k+1} n. \tag{20}$$

We first derive a formula for $p(E^i)$,[7] the probability of the event $v_i = 1$. We shall discard the two restrictions that *a*) all terms of U are independent, and b) that they are statistically identical. They need merely have continuous density functions. The formula is

$$p(E^i) = \sum_{j=i-k+1}^{i} p(R_j) - \sum_{j}^{i-1} \sum_{i-k+1} p(R_j R_{j+1}) \tag{21}$$

where R_j is the event that u_i is the largest variable in the sequence $u_j, u_{j+1}, \cdots, u_{j+k-1}$, $j = i - k + 1, i - k + 2, \cdots, i$. By the fundamental theorem on the probability of the union of events[8] we can write

$$p(E^i) = S_1 - S_2 + S_3 - \cdots \pm S_k,$$

where

$$S_1 = \sum_j p(R_j), \qquad S_2 = \sum_{j<l} p(R_j R_l),$$

$$S_3 = \sum_{j<l<m} p(R_j R_l R_m), \quad \text{etc.},$$

the sums running from $i - k + 1$ to i. Now in our problem all probabilities are determined exclusively by the first and last indexes, *e.g.*, $p(R_1 R_4) = p(R_1 R_3 R_4) = p(R_1 R_2 R_3 R_4)$. Since $\binom{n-1}{r}$ distinct product sets can be formed when there are r indexes between the first index and the last index $j + n$, and since

$$\sum_{r=0}^{n-1} (-1)^{r+1} \binom{n-1}{r} = 0 \qquad n \geq 2,$$

we arrive at (21).

[7] This derivation is due to L. Maximon "Some Properties of Infinite Sequences of Independent Random Variables," Group Rep. Nos. 34–53, Lincoln Lab., Lexington, Mass.; July 11, 1956.

[8] O. O. Feller, "Probability Theory and Its Applications," John Wiley and Sons, New York, N. Y., p. 61; 1950.

In the special case that all u_i are independent and identically distributed, we have

$$p(R_j) = \frac{1}{k} \quad \text{and} \quad p(R_j R_{j+1}) = \frac{1}{k+1}$$

for all j. Hence by (21), the probability that $v_i = 1$ is

$$k \frac{1}{k} - (k-1) \frac{1}{k+1} = \frac{2}{k+1},$$

and so the probability that $v_i = 0$ is

$$1 - \frac{2}{k+1} = \frac{k-1}{k+1}.$$

Eq. (20) then follows.

Alternate Recursion Formula for $p_n^2(r)$[9]

We have derived another recursion formula for $p_n^2(r)$:

$$p_n^2(r) = \frac{2r}{n} p_{n-1}^2(r) + \frac{n - 2(r-1)}{n} p_{n-1}^2(r-1) \quad \begin{matrix} n \geq 1 \\ r \geq 1 \end{matrix}. \tag{22}$$

The proof is quite simple. Consider a sequence X_1 of integers from 2 to n. Given that all x_i are statistically identical, if the integer 1 is inserted into the sequence to form a new sequence X_2, this integer can appear with equal probability at any place i in X_2. If Z_1 has $r - 1$ zeros, Z_2 must have either $r - 1$ or r zeros, or, alternately, if Z_2 has r zeros, Z_1 must have had $r - 1$ or r zeros. If the integer 1 is added adjacent to (on either side of) a minimum point of X_1, no change occurs; otherwise one zero is added. If there were r zeros in Z_1, which occurs with probability $p_{n-1}^2(r)$, there is probability $2r/n$ of adding no additional zero, since, for $k = 2$, two successive zeros cannot occur. Similarly, for $r - 1$ zeros in Z_1 [probability $p_{n-1}^2(r-1)$] there is probability $[n - 2(r-1)]/n$ of adding a zero. Thus we get (22).

This recursion makes it easy to compute successive values of $p_n^2(r)$. It is also possible to derive (11) from (22).

For $k = 3$ it is possible to write an analogous recursion formula for the joint density function $p_n(r_1, r_2)$, where r_1 is the number of single zeros (those bounded by 1's, *i.e.*, 101) and r_2 is the number of *pairs* of zeros (1001). Since three successive zeros are impossible for $k = 3$, $r = r_1 + 2r_2$, and $p_n^3(r)$ can be found by summing the $p_n(r_1, r_2)$ over r_1 and r_2. We shall not reproduce this untidy formula here.

Appendix

We list $a_n(y)$, with k indicated at the top of each list. The values of $p_n^k(r)$ can be obtained by comparing directly with (4). For example, $p_6^2(3)$ is the coefficient of y^3 in $a_6(y)$ under $k = 2$. Thus, $p_6^2(3) = 17/45$.

[9] This result was suggested by A. Kohlenberg.

$k = 2$

$$a_0(y) = 1$$

$$a_1(y) = y$$

$$a_2(y) = y$$

$$a_3(y) = \frac{2y + y^2}{3}$$

$$a_4(y) = \frac{y + 2y^2}{3}$$

$$a_5(y) = \frac{2y + 11y^2 + 2y^3}{15}$$

$$a_6(y) = \frac{2y + 26y^2 + 17y^3}{45}$$

$$a_7(y) = \frac{4y + 114y^2 + 180y^3 + 17y^4}{315}$$

$$a_8(y) = \frac{y + 60y^2 + 192y^3 + 62y^4}{315}$$

$$a_9(y) = \frac{2y + 247y^2 + 1452y^3 + 1072^4 + 62y^5}{2835}$$

$$a_{10}(y) = \frac{2y + 502y^2 + 5097y^3 + 7192y^4 + 1382y^5}{14175}$$

$$a_{11}(y) = \frac{4y + 2026y^2 + 34096y^3 + 83021y^4 + 35396y^5 + 1382y^6}{155925}$$

$$a_{12}(y) = \frac{2y + 2036y^2 + 55196y^3 + 217186y^4 + 171511y^5 + 21844y^6}{467775}$$

$$a_{13}(y) = \frac{4y + 8166y^2 + 349500y^3 + 2123860y^4 + 2801040y^5 + 776661y^6 + 21844y^7}{6081075}$$

$k = 3$

$$a_0(y) = 1$$

$$a_1(y) = y$$

$$a_2(y) = y^2$$

$$a_3(y) = y^2$$

$$a_4(y) = \frac{y^2 + y^3}{2}$$

$$a_5(y) = \frac{y^2 + 3y^3 + y^4}{5}$$

$$a_6(y) = \frac{2y^2 + 11y^3 + 17y^4}{30}$$

$$a_7(y) = \frac{2y^2 + 17y^3 + 65y^4 + 21y^5}{105}$$

$$a_8(y) = \frac{2y^2 + 24y^3 + 177y^4 + 196y^5 + 21y^6}{420}$$

$$a_9(y) = \frac{4y^2 + 64y^3 + 819y^4 + 1934y^5 + 959y^6}{3780}$$

$$a_{10}(y) = \frac{4y^2 + 82y^3 + 1737y^4 + 7037y^5 + 8717y^6 + 1323y^7}{18900}$$

$$a_{11}(y) = \frac{4y^2 + 102y^3 + 3515y^4 + 21635y^5 + 51030y^6 + 26341y^7 + 1323y^8}{103950}$$

$$a_{12}(y) = \frac{8y^2 + 248y^3 + 13894y^4 + 120149y^5 + 465379y^6 + 523031y^7 + 124691y^8}{1247400}$$

$$a_{13}(y) = \frac{16y^2 + 592y^3 + 54392y^4 + 625222y^5 + 3639482y^6 + 7209496y^7 + 4318450y^8 + 368550y^9}{16216200}$$

$k = 4$

$$a_0(y) = 1$$

$$a_1(y) = y$$

$$a_2(y) = y^2$$

$$a_3(y) = y^3$$

$$a_4(y) = y^3$$

$$a_5(y) = \frac{2y^3 + 3y^4}{5}$$

$$a_6(y) = \frac{2y^3 + 8y^4 + 5y^5}{15}$$

$$a_7(y) = \frac{4y^3 + 28y^4 + 58y^5 + 15y}{105}$$

$$a_8(y) = \frac{2y^3 + 21y^4 + 78y^5 + 109y^6}{210}$$

$$a_9(y) = \frac{2y^3 + 29y^4 + 162y^5 + 526y^6 + 226y^7}{945}$$

$$a_{10}(y) = \frac{4y^3 + 76y^4 + 585y^5 + 3266y^6 + 4619y^7 + 900y^8}{9450}$$

$$a_{11}(y) = \frac{8y^3 + 192y^4 + 1930y^5 + 16414y^6 + 46434y^7 + 36272y^8 + 2700y^9}{103950}$$

$$a_{12}(y) = \frac{8y^3 + 236y^4 + 2986y^5 + 36533y^6 + 168787y^7 + 288095y^8 + 127055y^9}{623700}$$

$$a_{13}(y) = \frac{16y^3 + 568y^4 + 8804y^5 + 150874y^6 + 1025054y^7 + 2949718y^8 + 3347326y^9 + 625740y^{10}}{8108100}$$

$k = 5$

$$a_0(y) = 1$$

$$a_1(y) = y$$

$$a_2(y) = y^2$$

$$a_3(y) = y^3$$

$$a_4(y) = y^4$$

$$a_5(y) = y^4$$

$$a_6(y) = \frac{y^4 + 2y^5}{3}$$

$$a_7(y) = \frac{2y^4 + 10y^5 + 9y^6}{21}$$

$$a_8(y) = \frac{2y^4 + 17y^5 + 44y^6 + 21y^7}{84}$$

$$a_9(y) = \frac{2y^4 + 25y^5 + 112y^6 + 197y^7 + 42y^8}{378}$$

$$a_{10}(y) = \frac{4y^4 + 68y^5 + 449y^6 + 1402y^7 + 1857y^8}{3780}$$

$$a_{11}(y) = \frac{4y^4 + 88y^5 + 789y^6 + 3647y^7 + 10757y^8 + 5505y^9}{20790}$$

$$a_{12}(y) = \frac{8y^4 + 220y^5 + 2546y^6 + 15973y^7 + 75491y^8 + 122407y^9 + 32835y^{10}}{249480}$$

$$a_{13}(y) = \frac{16y^4 + 536y^5 + 7732y^6 + 62498y^7 + 417898y^8 + 1285346y^9 + 1289034y^{10} + 180180y^{11}}{3243240}$$

$k = 6$

$$a_0(y) = 1$$

$$a_1(y) = y$$

$$a_2(y) = y^2$$

$$a_3(y) = y^3$$

$$a_4(y) = y^4$$

$$a_5(y) = y^5$$

$$a_6(y) = y^5$$

$$a_7(y) = \frac{2y^5 + 5y^6}{7}$$

$$a_8(y) = \frac{y^5 + 6y^6 + 7y^7}{14}$$

$$a_9(y) = \frac{y^5 + 10y^6 + 31y^7 + 21y^8}{63}$$

$$a_{10}(y) = \frac{2y^5 + 29y^6 + 152y^7 + 321y^8 + 126y^9}{630}$$

$$a_{11}(y) = \frac{2y^5 + 39y^6 + 297y^7 + 1081y^8 + 1731y^9 + 315y^{10}}{3465}$$

$$a_{12}(y) = \frac{4y^5 + 100y^6 + 1023y^7 + 5429y^8 + 15353y^9 + 19671y^{10}}{41580}$$

$$a_{13}(y) = \frac{4y^5 + 124y^6 + 1623y^7 + 11567y^8 + 47927y^9 + 132579y^{10} + 76446y^{11}}{270270}$$

Correspondence

Two Properties of Pseudo-Random Sequences*

The so-called "pseudo-random" (p-r) sequences defined below have two properties which may make them useful in certain communication systems. Stated roughly, the first property is that the minimum distance between different signals of a special class is maximized and the second is that the probability that one member of this class be mistaken for some other member is minimized.

Let $A_0 = (a_1, a_2, \cdots, a_n)$ be a sequence of length n in which $a_i = +1$ or $a_i = -1$ and let A_k be its k-th cyclic permutation. That is, $A_k = (a_{k+1}, a_{k+2}, \cdots, a_n, a_1, \cdots, a_k)$ for $k = 0, 1, \cdots, n-1$. In analogy with Slepian's[1] terminology, the set $A_0, A_1, \cdots, A_{n-1}$ shall be called an n-digit permutation alphabet if no two of the sequences A_i are identical. The transmission of these sequences over a symmetric binary channel in which the probability of erroneous reception of any digit a_i is p, independent of errors in the other digits, will be the concern here. The detection of such sequences may be considered to be a form of digital phase detection. Barker[2] and Sherman[3] have considered similar problems.

Let r_j be the covariance coefficient, defined by

$$r_j = \langle A_k A_{k+j} \rangle = \sum_{i=1}^{n} a_{k+i} a_{k+i+j}, \quad (1)$$

where $a_{i+n} = a_i$ and $A_{i+n} = A_i$. Clearly,

$$r_j = \sum_{i=1}^{n} a_i a_{i+j}, \quad (2)$$

for any value of k in (1). The sequences A_i will be called "pseudo-random" if the covariance coefficients satisfy the relations

$$r_0 = n$$

$$r_j = -1 \quad (j = 1, 2, \cdots, n-1). \quad (3)$$

The alphabet $A_0, \cdots, A_{n-1}$ will be called an n-digit p-r alphabet if (3) holds. The name "pseudo-random" has been used by some authors[4] because of the resemblance of these sequences to white noise, as shown by (3). It should be noted that p-r sequences do not exist for all values of n. In fact, it can easily be shown[4] that a necessary (but not sufficient) condition for the existence of a p-r sequence is that n have the form $4m - 1$, where m is an integer. Note that (2) and (3) do not specify the n quantities a_i uniquely since the covariance coefficients are not independent. In fact, it is easily shown from (2) that $r_j = r_{n-j}$ for all j.

Huffman's[5] "maximum length null sequences," all of which are of length $2^m - 1$ (m an integer), have the p-r property [see (3)]. Barker[6] has given sequences of length 3, 7 and 11 which satisfy (3). Henceforth, it will be assumed that there exists at least one p-r alphabet of the given length n. We will show that this alphabet is the best by two different criteria of all n-digit permutation alphabets.

First, it will be shown that the minimum distance between different sequences of a permutation alphabet is maximized by choosing a p-r alphabet. As usual, the distance between two sequences is defined as the number of digits which must be altered to change the first sequence into the second. Since the product $(a_{k+i})(a_{k+i+j})$ has the value $+1$ if $a_{k+i} = a_{k+i+j}$ and otherwise has the value -1, it is easily seen that the distance between A_k and A_{k+j} is $\frac{1}{2}(n - r_j)$. Let D be the minimum distance between any two sequences of an n-digit permutation alphabet. That is,

$$D = \min_{1 \le j \le n-1} \tfrac{1}{2}(n - r_j). \quad (4)$$

For a p-r alphabet it follows from (3) that the minimum distance, D_p, is given by

$$D_p = \tfrac{1}{2}(n + 1). \quad (5)$$

The following property will be proven:

Theorem 1) The minimum distance for an n-digit pseudo-random alphabet is greater than or equal to the minimum distance for any other n-digit permutation alphabet.

Proof: It is easily shown from (2) that

$$\sum_{j=1}^{n-1} r_j = c^2 - n, \quad (6)$$

and hence that

$$\sum_{j=1}^{n-1} (n - r_j) = n^2 - c^2, \quad (7)$$

where

$$c = \sum_{j=1}^{n} a_j. \quad (8)$$

Therefore, since $2D \le n - r_j$,

$$2(n-1)D \le n^2 - c^2 \le n^2 - 1. \quad (9)$$

The last step follows from the observation that $c^2 \ge 1$ for odd n. Since a p-r sequence of length n exists, n is necessarily odd. Therefore, from (9) and (5),

$$D \le \tfrac{1}{2}(n+1) = D_p. \quad (10)$$

This completes the proof.

The minimum distance is one criterion which might be used in evaluating error-correcting or other codes. Huffman[7] has also discussed the error-correcting properties of maximum length null sequences from another point of view.

The second property of p-r alphabets to be proven may be applied to systems in which there is error detection but no error correction. In such systems the probability of an incorrect decision is minimized by the use of p-r alphabets.

Theorem 2) The probability that a transmitted sequence of an n-digit permutation alphabet will be incorrectly identified as some other sequence of the alphabet is greater than or equal to the corresponding probability for an n-digit pseudo-random alphabet.

Proof: Since the number of digits which must be altered to change A_k to A_{k+j} is $\frac{1}{2}(n - r_j)$, the probability that A_k will be identified by the receiver as some other sequence is

$$P = \sum_{j=1}^{n-1} p^{\frac{1}{2}(n-r_j)} (1-p)^{\frac{1}{2}(n+r_j)}. \quad (11)$$

The values r_j which minimize P when these values are subjected to the constraint (6) are now determined. This problem is solved easily by the method of Lagrange multipliers.[8] The result is that one must have

$$r_j = -(n - c^2)/(n-1), \quad (j = 1, 2, \cdots, n-1) \quad (12)$$

When all the coefficients r_j are equal, it is clear from (11) that when $p < \frac{1}{2}$ (which is assumed), P is minimized by choosing these coefficients to be as small as possible. Since $c^2 \ge 1$, the minimum is attained when $r_j = -1$ ($j = 1, 2, \cdots, n-1$). Thus, P is least when $r_j = -1$ ($j = 1, 2, \cdots, n-1$), *i.e.*, when $A_0, \cdots, A_{n-1}$ is a p-r alphabet.

L. Lorne Campbell
Defence Res. Telecommun. Establ.
Ottawa, Ont. Can.

* Received by the PGIT, June 5, 1958. This work was performed under project PCC No. D48-28-01-07.

[1] D. Slepian, "A class of binary signaling alphabets," *Bell. Sys. Tech. J.*, vol. 35, pp. 203–234; January, 1956.

[2] R. H. Barker, "Group synchronizing of binary digital systems," in "Communication Theory," W. Jackson (Ed.), Butterworths Scientific Publications, London, Eng., pp. 273–287; 1953.

[3] H. Sherman, "Some optimal signals for time measurement," IRE Trans. on Information Theory, vol. IT-2, pp. 24–28; March, 1956.

[4] D. W. Lytle, "On the properties of matched filters," Stanford Electronics Lab., Stanford Univ., Menlo Park, Calif., Tech. Rep. No. 17; June, 1957.

[5] D. A. Huffman, "The synthesis of linear sequential coding networks," in "Information Theory, Third London Symposium," C. Cherry, (ed.) Butterworths Scientific Publications, London, Eng., pp. 71–95; 1956. See particularly p. 91.

[6] Barker, *op. cit.*, p. 282.

[7] D. A. Huffman, "A linear circuit viewpoint on error-correcting codes," IRE Trans. on Information Theory, vol. IT-2, pp. 20–28; September, 1956.

[8] R. Courant and D. Hilbert, "Methods of Mathematical Physics," Interscience Publishers, Inc., New York, N. Y., vol. 1, p. 165; 1953.

n Inequality Concerning the nvelope of a Correlation unction*

The following inequality for a correlation nction $R(\tau)$ is proved:

$$R(0) \geq [R^2(\tau) + I^2(\tau)]^{1/2} \quad (1)$$

here, in terms of a "power spectrum" or ectral density $w(f)$,

$$R(\tau) \equiv \int_0^\infty w(f) \cos 2\pi f\tau \, df \quad (2)$$

nd

$$I(\tau) \equiv \int_0^\infty w(f) \sin 2\pi f\tau \, df. \quad (3)$$

This inequality, which is a stronger atement than the well-known fact that $(0) \geq |R(\tau)|$, is equivalent to the state- ent (hereby proven) that the envelope roperly defined) of a correlation function a maximum at the origin.

To prove (1), consider the following uantity, which may be thought of as a mplex version of $R(\tau)$:

$$\int_0^\infty w(f)e^{2\pi i f\tau} \, df \equiv Q(\tau)e^{i\theta(\tau)}. \quad (4)$$

Considering the integral in (4) and membering that $w(f)$ is non-negative, we ay write,

$$\left| \int_0^\infty w(f)e^{2\pi i f\tau} \, df \right| \leq \int_0^\infty w(f) \, df \quad (5)$$

nce the absolute value of the integral ust be no greater than the integral of e absolute value of the integrand. The ft-hand side of (5) is, however,

$$[R^2(\tau) + I^2(\tau)]^{1/2}$$

hereas the right-hand side is $R(0)$. Thus e inequality (1) is proved.

To demonstrate the connection of the equality with the envelope of $R(\tau)$, the velope we define in the following way:[1] onsider the definition (4), in which $Q(\tau)$ nd $\theta(\tau)$ are taken to be real. Equating al and imaginary parts of (4) we obtain

$$Q(\tau) \cos [\theta(\tau)] = R(\tau) \quad (6)$$

nd

$$Q(\tau) \sin [\theta(\tau)] = I(\tau). \quad (7)$$

'e obtain, further,

$$Q(\tau) = [R^2(\tau) + I^2(\tau)]^{1/2} \quad (8)$$

where the positive square root is used. Then $\cos [\theta(\tau)]$ and $\sin [\theta(\tau)]$ are unambiguously defined, from (6) and (7). It can now be seen by (6) that a correlation function $R(\tau)$ can be written as the product of $Q(\tau)$ and $\cos [\theta(\tau)]$, where $Q(\tau)$ and $\cos [\theta(\tau)]$ are unambiguously defined and where $|\cos [\theta(\tau)]| \leq 1$. $Q(\tau)$ is now called the envelope of $R(\tau)$. This cognomen is seen to be natural and convincing. One may verify by specific calculation that the envelope as defined by (8) coincides, at least in many cases, with what one would consider the "coarse" structure of $R(\tau)$ As an example of the fundamental nature of this envelope $Q(\tau)$, one may point to a recently published paper,[2] in which it is shown that the output of a "band-pass correlation detector" is what we have called $Q(\tau)$.

Since $Q(0) = R(0)$, the inequality (1) is now seen to state that $Q(0) \geq Q(\tau)$, *i.e.*, the envelope of a correlation function has a maximum at the origin of τ. This is stronger than the statement that $R(0) \geq |R(\tau)|$ since $I^2(\tau)$ is non-negative.

Philip R. Karr
Ramo Wooldridge
Div. of Thompson Ramo-Wooldridge Inc.
Los Angeles, Calif.

* Received by the PGIT, September 22, 1958.

[1] The envelope of $R(\tau)$ is defined with the aid its Hilbert transform $I(\tau)$. The use of Hilbert ansforms for similar purposes has been invoked a number of authors. See for example, J. Dug- dji, "Envelope and pre-envelopes of real wave- rms," IRE Trans. on Information Theory, l. IT-9, pp. 53–57; March, 1958, where a number other references are cited.

[2] P. E. Green Jr. "The output signal-to-noise ratio of correlation detectors," IRE Trans. on Information Theory, vol. IT-3, pp. 10–18; March, 1957.

Lossless Symbol Coding with Nonprimes*

Golay[1] has asked the question as to "whether the search for master iteration matrices, of the form p^n for values of n higher than 2 can be systematized." There exists a straight-forward method for constructing the coding matrix for realizing all the one-error correction codes whose existence was shown by Zaremba.[2] The ability to do this depends upon the fact that there exist finite fields of order p^n, where p is a prime, namely the Galois fields. The method is completely constructive since in order to construct the Galois field $GF(p^n)$ it is only necessary to find a polynomial of degree n which is irreducible over the field of integers mod p. This is constructive since there is only a finite number of these polynomials; if one can't find a better way to obtain an irreducible polynomial, one can test each of these polynomials for irreducibility. Assuming that a Galois field of order p^n has been constructed, the following method is one way of obtaining the coding matrix. Incidentally, the method proves constructively the theorem of Zaremba as to the existence of such codes.

Let $0, 1, a_3, \cdots, a_{p^n}$ be the elements of a Galois field $GF(p^n)$ and H be the prime ideal defining said field. Suppose one wishes to construct a single-symbol correcting code with k check symbols. First write down all possible distinct column vectors v_i of length k deleting the vector of the form 0, and, although this is not necessary, make the first k vectors of the list of the form $v_i = \delta_{ij}$ (kronecker delta); this simplifies encoding and decoding. One has $p^{nk} - 1$ distinct vectors. Now systematically delete from the list all vectors which are of the form $v_j = av_i$ for $i < j$ and $a \in GF(p^n)$. Since a field has no zero divisors, one clearly has left $p^{nk} - 1/p^n - 1$ vectors. Now construct a code with $s = p^{nk} - 1/p^n - 1$ total symbols, k check symbols, and $s - k$ information symbols. Let A be the matrix composed of the vectors left after the deletion process. The column order is the same as left by the processes.

Given $s - k$ information symbols, form a column vector w of length s, the first k elements from the top being 0 and the remaining $s - k$ elements composing the given information symbols. Forming $Aw \equiv z \bmod (H)$, one obtains a column vector of length k. Since the elements of the vector are in a field, the additive inverse $-z$ of z exists. Form a vector x which has $-z$ as its first k elements and the information symbols as the remaining $s - k$ elements. Since the first k columns of the matrix are of the form $v_i = \delta_{ij}$, we see that $Ax \equiv 0 \bmod H$. This is now the encoded message. Suppose there is an error of amount b in the eth character. Then the received message is $x + \delta_{ej}b$. Form $A(x + \delta_{ej}b) \equiv A\delta_{ej}b \equiv v_eb \bmod (H)$.

Looking at v_eb we can determine that the eth character must be in error, for by construction of A, v_eb is not of the form v_ja for any $j \neq e$ and $a \in GF(p^n)$. If this were so we would have $v_eb = v_ja$ or $v_e = v_jab^{-1}$, and v_e would be the multiple of a vector v_j; and this is not the case by construction. Thus we know that the eth character is in error and, further solving $v_ey = v_eb$, we can obtain $y = b$ and thus correct any single symbol error. It might be noted that this is as far as this particular method can be extended since we are relying on the property that the elements form a field and every finite field is isomorphic to some Galois field.

In order to determine the Master Iteration Matrices (MIM) so that one can check the codes using the multiplication table of the integers mod p rather than the multiplication table of the Galois field, the following procedure is easily seen to work. Each element of a field $GF(p^n)$ can be represented as a polynomial of the form $\sum_{i=0}^{n-1} c_ix^i$ where c_i are the integers mod p. Make the correspondence $c = \sum_{i=0}^{n-1} c_ix^i \leftrightarrow (c_0, c_1 \cdots c_{n-1}) = c'$ making this a column vector. If $a \in GF(p^n)$ is a typical element of the coding matrix A, form the product $a \cdot c$ where $a = \sum_{i=0}^{n-1} a_ix^i$ and the a_i are the particular coefficients in a, and $c =$

* Received by the PGIT, January 26, 1959.

[1] M. J. E. Golay, "Notes on the penny-weighting problem, lossless symbol coding with nonprimes, etc.," IRE Trans. on Information Theory, vol. IT-4, pp. 103–109; September, 1958.

[2] S. K. Zaremba, "Covering problems concerning Abelian groups," *J. London Math. Soc.*, vol. 27, pp. 242–246; April, 1952.

$\sum_{i=0}^{n-1} c_i x^i$ where the c_i are indeterminates. Then using the polynomial H to reduce all the powers we determine a matrix a^* which if

$$c = \sum_{i=0}^{n} c_i x^i \leftrightarrow (c_0 c_1 \cdots c_{n-1}) = c'$$

and

$$ac = \sum a_i x^i \sum c_i x^i$$

$$\leftrightarrow (d_0 \cdots d_{n-1}) = d'$$

then $a^* c' = d'$. Thus matrix a^* is the MIM searched for. Replacing the elements of A by the corresponding MIM we obtain a new matrix A' which operates on vectors where each element of the message is represented by a column vector of the form $(c_0 \cdots c_{n-1})$ where the c_i's are integers mod p. This matrix is used to encode the message as before but the congruence is mod p instead of mod H.

Example: Let 0, 1, x, $x + 1$ be the elements of $GF(2^2)$ where $H = x^2 + x + 1$. Then let $k = 2$; we calculate $s = 5$; a possible matrix turns out to be:

$$\begin{pmatrix} 1 & 0 & 1 & 1 & x+1 \\ 0 & 1 & 1 & x+1 & 1 \end{pmatrix}.$$

To form the MIM we see that $1 \cdot (c_1 x + c_0) = c_1 x + c_0$, so 1 gives $\binom{10}{01}$; and $(x + 1)(c_1 x + c_0) = c_1(x^2 + x) + c_x + c_0 = (xc_0) + c_0 + c_1$, so $(x + 1)$ gives $\binom{11}{10}$. Thus the matrix becomes

$$\begin{array}{cccccccccc} 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 1 \end{array}.$$

One thing that is interesting is that the Hamming Code has lossless points at $s = 2^i - 1 = 1, 3, 7, 15, 31, 63 \ldots$. For the lossless codes of radix 2^2 we obtain $s = 4^k 1/3 = 1, 5, 13 \ldots$ or, since each character is 2 bits, it is bitwise 1, 26, ... and the number of check bits would be 2, 4, 6. Comparing this with the Hamming code, notice for instance that for $s = 10$ one must use 4 bits so that what this code does to become lossless is to correct 5 double errors. Thus for radix 2^2 and assuming that the errors are not correlated and that all have equal probability we are able to construct some best possible codes. This, of course, can be done easily by the proper interpretation of the Hamming check bits. However, it interested me that this code did it automatically.

John Cocke
IBM
Poughkeepsie, N. Y.

A Comment on a Comment on Pattern Redundancy*

In a recent letter[1] concerning an article by Glovazky,[2] Lowenschuss gives a set of necessary constraints on the choice of a code schedule for distinguishing overspecified black and white patterns. These constraints are in fact not necessary except in the case that $P = 2^p$. This can be seen from the following simple counter example. We have nine patterns specified by the four bit codes given in Fig. 1. In the terminology of Lowenschuss this would make $C = 4$, $P = 9$. The constraints derived using Lowenschuss' necessary conditions are

$$4 \leq c_0 \leq 5$$

$$4 \leq c_1 \leq 5$$

$$4 \leq c_2 \leq 5$$

$$1 \leq c_3 \leq 8$$

Or, in other words, three of the columns chosen should have either four or five "ones." Examination of the number of ones in the columns of Fig. 1 reveals that there is only one column which meets this condition. According to Lowenschuss, separation with four cells would then be impossible. However, the four cells given do result in separation (no two of the rows of Fig. 1 are the same).

Patterns \ Columns	1	2	3	4
1	0	1	1	0
2	0	1	1	1
3	1	0	0	1
4	1	0	1	0
5	1	0	1	1
6	1	1	0	0
7	1	1	0	1
8	1	1	1	0
9	1	1	1	1
Column Sums—	7	6	6	5

Fig.—1

Actually Lowenschuss' necessary condition is applicable if separation must be obtained into consecutive codes. That is, if the only possible correct choices of cells are those which yield a set of binary codes ordered according to their numerical value. This condition, however, is not necessary for separation.

One can obtain a generalization of the constraint given by Lowenschuss for $P = 2^p$. The procedure is as follows: Suppose we are given P codes and are attempting to affect separation with C cells. Then we can derive a constraint on the total number of ones in all C cell columns. All 2^C C-bit binary numbers are listed; first the binary number having C ones, followed by all those binary numbers having $C - 1$ ones, then all those having $C - 2$ ones and so forth. It follows that if we count the number of ones in the first P binary numbers of this listing, the result will be an upper bound on the total number of ones allowable in all C cell columns. Similarly we can derive a constraint on the total number of ones in any $C - 1$ of the n columns. All 2^C C-bit binary numbers are listed; first all the binary numbers with $C - 1$ ones in the first $C - 1$ columns, then all those having $C - 2$ ones in the first $C - 1$ columns and so forth. It follows that if we count the total number of ones in the first $C - 1$ columns of the first P binary numbers of this listing, we have an upper bound on the total number of ones in any $C - 1$ of the C cell columns chosen for separation.

Similarly upper bounds can be derived on the number of ones in any k columns ($k = 1, 2, \cdots C$). The calculation of these upper bounds is summarized symbolically below.

$\binom{k}{i}$ = Number of combinations of "k" things taken "i" at a time.
P = Number of codes.
C = Number of separation cells.
T^k = Upper bound on the number of ones in any k of the C separation cells.

$$T^k = 2^{C-k} \sum_{i=0}^{i=k} f_i^k(i) \binom{k}{i}$$

where
f_k^k = Minimum of

$$\left[1, \text{ and } \frac{P}{\binom{k}{k} 2^{C-k}}\right]$$

and

f_i^k = Minimum of

$$\left[1, \text{ and } \frac{P - 2^{C-k} \sum_{j=i+1}^{j=k} f_i^k \binom{k}{j}}{2^{C-k} \binom{k}{i}}\right]$$

for $k \neq i$.

As an example let $P = 11$, and $C = 4$. Then we find that

$T^1 = 8 \geq$ number of ones in any single column
$T^2 = 15 \geq$ number of ones in any two columns
$T^3 = 21 \geq$ number of ones in any three columns, and
$T^4 = 28 \geq$ number of ones in all four columns.

* Received by the PGIT, January 26, 1959.

[1] A. Glovazky, "Determination of redundancies in a set of patterns," IRE Trans. on Information Theory, vol. IT-2, pp. 151–153; December, 1956.

[2] O. Lowenschuss, "A comment on pattern redundancy," IRE Trans. on Information Theory, vol. IT-4, p. 127; September, 1958.

Now we can get constraints on individual umns. If c_i is the number of ones in the cell column, and we agree to order the so that if $i < j$ then $c_i \geq c_j$ we can rite the above constraints as follows:

$$c_1 \leq 8$$

$$c_1 + c_2 \leq 15$$

$$c_1 + c_2 + c_3 \leq 21$$

$$c_1 + c_2 + c_3 + c_4 \leq 28.$$

lving these for the individual c_i we obtain

$$c_1 \leq 8$$

$$c_2 \leq 7$$

$$c_3 \leq 7$$

$$c_4 \leq 7.$$

In general we can express these individual column constraints as

$$c_i \leq \left[\text{minimum of } \left(\frac{T^1}{1}, \frac{T^2}{2}, \cdots \frac{T^i}{i} \right) \right]$$

where [ ] means the next lowest integer. Furthermore it can be shown that $T^j/j \leq T^{j-1}/j - 1$, so we have finally

$$c_i \leq \left[\frac{T^i}{i} \right].$$

The upper bounds on the number of zeros in the C cell columns are the same as those for the number of ones, and can be used to get lower bounds on the number of ones in the cell columns.

The calculation of the constraints described here is not a trivial matter, but calculation would seem worthwhile if efficiency was of concern in determining a code schedule.

MARVIN C. PAULL
Bell Telephone Labs.,
Whippany, N. J.

Contributors

Marshall Freimer (M '56) was born in w York, N. Y., on May 6, 1932. He eived the A.B. degree in 1953, and the A.M. degree in 1954, both in mathematics, from Harvard University, Cambridge, Mass. He is preparing a thesis for the Ph.D. degree in statistics, also at Harvard University.

M. FREIMER

Since June, 1957, he has been a staff member of the Lincoln Laboratory, assachusetts Institute of Technology, xington, Mass., working on combina- ial mathematics, dynamic programming, d adaptive processes.

Mr. Freimer is a member of Phi Beta ppa, Sigma Xi, the Institute of Mathe- atical Statistics, and the Mathematical sociation of America.

❖

Alan L. Tritter (M '56) was born in ston, Mass., on March 25, 1935. He eived the B.A. degree from the University of Chicago, Ill., in 1952. He did graduate work in mathematics and law from 1952–1955.

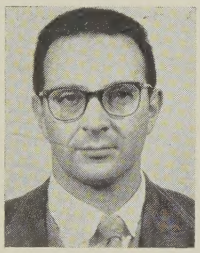
A. L. TRITTER

Since July, 1955, he has been a staff member at the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass., working in the fields of combinatorial mathematics, pattern recognition, and effective computer usage.

Mr. Tritter is a member of the American Mathematical Society, the Mathematical Association of America, and Gamma Alpha Graduate Scientific Fraternity.

❖

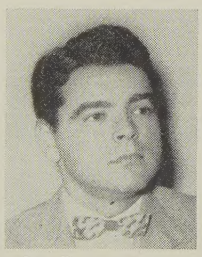
Bernard Gold was born in New York, N. Y., on March 31, 1923. He received the D.E.E. degree from the Brooklyn Polytechnic Institute, Brooklyn, N. Y., in 1948. Since then he has been with the Hughes Aircraft Company, Culver City, Calif., and the Lincoln Laboratory, Massachusetts Institute of Technology, Lexington, Mass., working in the fields of radar, communications, and pattern recognition.



B. GOLD

❖

Melvin J. Jacobson was born in Providence, R. I., on November 25, 1928. He received the A.B. degree from Brown University in 1950, and the M. S. degree and the Ph.D. degree in mathematics from the Carnegie Institute of Technology in 1952 and 1954, respectively.



M. J. JACOBSON

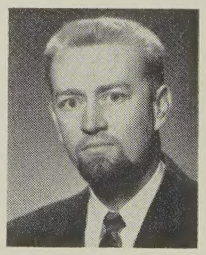
From 1952 to 1954 he was an instructor in mathematics and project mathematician at Carnegie, where he worked in the general area of fluid mechanics and lubrication. From 1954 to 1956 he was a member of the technical staff at Bell Telephone Laboratories, Inc., Whippany, N. J., where he was engaged in the solution of systems analysis problems. He was an assistant professor of mathematics at Rensselaer Polytechnic Institute, Troy, N. Y., from 1956 to 1958, where he is now an associate professor.

In addition to having published papers and given talks on the mathematical analysis of signal processing systems, he has published papers on the theory of lubrication of journal bearings. At the present time he is senior investigator of an Office of Naval Research Contract, under which receiving systems of the correlation type are being analyzed.

Dr. Jacobson is a member of the American Mathematical Society, the Acoustical Society of America, Sigma Xi, Phi Kappa Phi, and Pi Mu Epsilon.

❖

Kenneth J. Hammerle was born in Batesville, Ind., on December 15, 1922. He received the B.S.E.E. degree in 1945, the M.S.E.E. in 1947, and the Ph.D. degree in electrical engineering in 1951, all from Purdue University, Lafayette, Ind. He was a member of the teaching staff of the Department of Electrical Engineering at Purdue from 1946 to 1951.



K. J. HAMMERLE

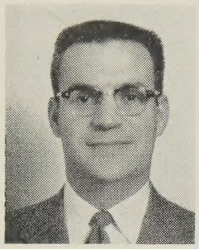
Since 1951 he has been employed by the Boeing Airplane Company, Seattle, Wash., where he has been

occupied with a variety of problems in radar and communications. He has also served on the electrical engineering staff of the Evening Division of Seattle University.

Dr. Hammerle is a member of Eta Kappa Nu, Tau Beta Pi, Sigma Xi, and Sigma Pi Sigma.

❖

Robert L. Brock (A'44—M'57) was born in Waverly, Kans., on November 28, 1923. He received the B.S. and M.S. degrees in physics in 1948 and 1950, from Oregon State College, Corvallis, where he was also a teaching assistant in 1948–1949. He did graduate work at both Washington and Wisconsin Universities, St. Louis, Mo., and Madison, Wis., respectively, serving as a teaching assistant in 1949–1950 at the latter. He received the Ph.D. degree in mathematics in 1955 from Oregon State College.



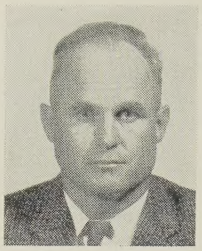
R. L. Brock

From 1942 to 1945 he served with the U. S. Navy as an aircraft radar technician and instructor with the early G.C.A. Blind Landing Systems. From 1951 to 1953 he was a research engineer at the Boeing Airplane Company, Seattle, Wash. He was instructor in mathematics at Oregon State College from 1953 to 1955, when he returned to Boeing as a research specialist in the Physical Research Staff, later joining the mathematics staff of the Boeing Scientific Research Laboratories. At present he is supervisor of the Mathematics Unit of the Boeing Systems Management Office, Seattle.

His principal research effort has been in the application of mathematics to electronics, including guidance and communication, statistical detection studies, information theory and problems in electromagnetic theory.

Dr. Brock is a member of Sigma Pi Sigma, Pi Mu Epsilon, Sigma Xi, Phi Kappa Phi, the American Mathematical Society, the Mathematical Association of America, and the Society for Industrial and Applied Mathematics.

William M. Stone was born on February 4, 1915, in Oregon City, Ore. In 1938 he received the B.A. degree from Willamette University, Salem, Ore., in 1940 the M.A. degree from Oregon State College, Corvallis, and in 1947 the Ph.D. degree from Iowa State College, Iowa City, all in mathematics.



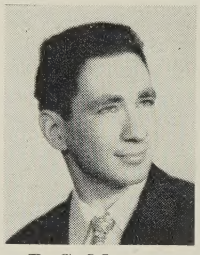
W. M. Stone

From 1947 to 1951 he served as assistant professor of mathematics at Oregon State College, and from 1951 to 1953 was employed as a research engineer at the Boeing Airplane Company, Seattle, Wash. Since 1953 he has been associate professor of mathematics at Oregon State, at the same time working in close conjunction with the Boeing Company on problems involving stochastic processes.

Dr. Stone is a member of the American Mathematical Society, the Mathematical Association of America, Sigma Xi, the American Association for the Advancement of Science, and the Society for Industrial and Applied Mathematics.

❖

Richard S. Marcus (S'54) was born in Atlantic City, N. J., on July 26, 1933. He received the A.B. and B.S. degrees in electrical engineering from the University of Pennsylvania, Philadelphia, in 1954 and 1955, respectively. His graduate work was done at the Massachusetts Institute of Technology, Cambridge, Mass., where he received the M.S. degree in 1957 and the E.E. degree in 1958.



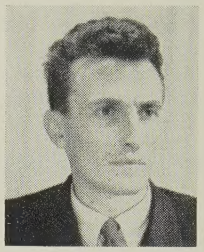
R. S. Marcus

In 1956 he became a research assistant in the Information Theory Group of the Research Laboratory of Electronics at M.I.T. In 1958 he joined the staff of the M.I.T. Servomechanisms Laboratory, where he now is working on intelligent computer-human systems.

Mr. Marcus is a member of Sigma Xi, Tau Beta Pi, and Eta Kappa Nu.

❖

M. P. Schützenberger was born on October 24, 1920, in Paris, France. He received the Ph.D. degree in mathematics from the University of Paris in 1950.



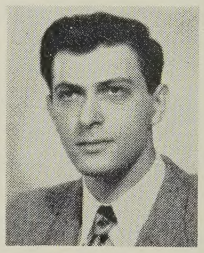
M. P. Schützenberger

He was a W.H.O. Consultant in Southeast Asia from 1950 to 1954, and a research assistant at the Massachusetts Institute of Technology, Cambridge, Mass., from 1955 to 1956. He is presently Maître de Conférences, Le Faculté des Sciences of Poitiers.

Dr. Schützenberger is a member of the International Institute of Statistics.

❖

Moshe Zakai (Zakhaim) was born in Sokolka, Poland, on December 22, 1926. He attended the Technion-Israel Institute of Technology, Haifa, Israel, where he received the B.S. degree in 1951 and the Ingenieur degree in 1952, both in electrical engineering.



M. Zakai

In 1952 he joined the Scientific Department, Ministry of Defense, Israel. He came to United States in 1956 on an Israeli Government fellowship, and received the Ph.D. degree in electrical engineering from the University of Illinois, Urbana, Ill., in 1958.

From February, 1958 to August, 1958, he was with the Electronics Research Laboratories, Columbia University, New York, N. Y., where he was engaged in systems design studies.

Dr. Zakai recently returned to Israel. He is now on the technical staff of the Scientific Department, Ministry of Defense.

INFORMATION FOR AUTHORS

Authors are requested to submit editorial correspondence or technical manuscripts to the Publications Chairman for possible publication in the PGIT TRANSACTIONS. Papers submitted should include a statement as to whether the material has been copyrighted, previously published, or accepted for publication elsewhere.

Papers should be written concisely, keeping to a minimum all introductory and historical material. It is seldom necessary to reproduce in their entirety previously published derivations, where a statement of results, with adequate references, will suffice.

To expedite reviewing procedures, it is requested that authors submit the original and two legible copies of all written and illustrative material. The manuscript should be double-spaced, and the illustrations drawn in India ink on drawing paper or drafting cloth. Each paper should include a carefully written abstract of not more than 200 words. Upon acceptance, papers should be prepared for publication in a manner similar to those intended for the PROCEEDINGS OF THE IRE. Further instructions may be obtained from the Publications Chairman. Material not accepted for publication will be returned.

IRE TRANSACTIONS ON INFORMATION THEORY is published four times a year, in March, June, September, and December. A minimum of one month must be allowed for review and correction of all accepted manuscripts. In addition, a period of approximately two months is required for the mechanical phases of publication and printing. Therefore, all manuscripts must be submitted three months prior to the respective publication dates.

All technical manuscripts and editorial correspondence should be addressed to George A. Deschamps, University of Illinois, Urbana, Ill. Local Chapter activities and announcements, as well as other nontechnical news items, should be addressed to Laurin G. Fischer, ITT Laboratories, 492 River Road, Nutley 10, N. J.

INSTITUTIONAL LISTINGS

The IRE Professional Group on Information Theory is grateful for the assistance given by the firms listed below and invites application for Institutional Listing from other firms interested in the field of Information Theory.

MOTOROLA, INC., 4545 West Augusta Blvd., Chicago 51, Ill.
Television, Home & Auto Radio, Phonograph & Hi-Fi, Communications & Industrial Electronics

THE RAMO-WOOLDRIDGE CORPORATION
5500 West El Segundo Blvd., Los Angeles 45, Calif.

REPUBLIC AVIATION CORP., Farmingdale, N. Y.
Aircraft, Missiles, Drones, Electronic Analyzers; U. S. Distr. of Alouette Turbine-Powered Helicopter